THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 58, PART 4 1 July 1946 No. 328

LIGHT EMISSION DURING COOLING OF A PLANCKIAN RADIATOR

By E. F. CALDIN,

Department of Physical and Inorganic Chemistry, Leeds University

MS. received 7 *December* 1945 ; *read* 15 *March* 1946

ABSTRACT. General expressions are derived for the variation of luminous intensity with time, and the corresponding integral of the intensity with respect to time, for a Planckian radiator of uniform temperature, cooling by radiation from temperatures above 2000° K. The time-intensity integral for light emitted in a given direction is found to depend only on the initial temperature, heat capacity and shape of the body ; the form of the intensity-time curve depends also on the surface area and emissivity. Values of the time-intensity integral for a spherical total radiator of unit heat capacity, cooling from various temperatures, are tabulated. The relevance of these results to the study of certain light sources of short duration is indicated.

§ 1. INTRODUCTION

THE study of transient light sources of short duration presents special problems because of the rapid continuous variation of the luminous intensity of the source. Examples of such sources are: (*a*) metal wires burning in oxygen (Van Liempt and de Vriend, 1934, 1935); (*b*) photographic flash powders, such as the Agfa powder, consisting of a mixture of magnesium and thorium nitrate (Beck and Eggert, 1926); and (*c*) the photoflash bulb used in photography, such as the Sashalite bulb, in which aluminium foil burns in oxygen (Reeb, 1935; Van Liempt and de Vriend, 1937). The luminous intensity of such a source increases rapidly to a maximum and then exhibits a much slower decrease which is roughly exponential with time (Beck and Eggert, 1926; Reeb, 1935; Van Liempt and de Vriend, 1937; Kelley, 1938; Projector and Barbrow, 1945). These sources exhibit continuous spectra (cf. following paper) and may well consist essentially of incandescent solid particles—the oxide produced by combustion of the metal—raised to incandescence by the heat of combustion, and cooling by radiation. It is thus of interest to compare the declining part of the intensity-time curve with the corresponding curves for Planckian radiators, initially at a high temperature, and cooling by radiation. Further, the actinic effect of such a source, i.e. the density of a photographic plate exposed to it under given conditions, will depend (neglecting reciprocity-law failure) not on the maximum intensity but on the integral with respect to time of the intensity, i.e. the area under the intensity-time curve; so that the value of this integral for a Planckian radiator is also of interest. Experimental values for the time-intensity integral for the above sources have been given by Beck and Eggert (1926), Eder (1930), Reeb (1935), and Van Liempt and de Vriend (1934, 1935).

In this paper, general expressions are derived for the variation of luminous intensity with time, and for the integral of the intensity with respect to time, for Planckian radiators cooling from temperatures above 2000° K. The derivation makes use of the relation already reported between the brightness and temperature of a total radiator (Caldin, 1945). To obtain manageable expressions, the temperature of the cooling body has been assumed uniform and temperature gradients ignored. A table of values is given for the time-intensity integral for a spherical total radiator of unit heat capacity, cooling from various temperatures; this allows easy calculation of the value of the integral for other simple cases. The conditions necessary for the valid application of the theory to data on actual sources are being examined by Dr. F. R. Booth. None of the published experimental work is complete enough to allow a satisfactory test of the theory; some new experimental work will be presented, it is hoped, in a subsequent paper.

The question of units of intensity arises. Calculations using the Planck equation and the standard C.I.E. relative luminosity data give values of $\int_0^\infty V_\lambda J_\lambda d\lambda$. Here V_λ is the value of the C.I.E. relative luminosity factor at the wave-length λ; $J_\lambda d\lambda$ is the total energy flux from unit area of the total radiator between the wave-lengths $\lambda \pm d\lambda/2$, as given by Planck's equation. If $J_\lambda d\lambda$ is expressed in watts per cm.^{2}, the luminous flux per unit area is then given in lightwatts per cm.^{2}, according to the definition of Moon (1936). It would be very convenient if corresponding units were in use for intensity, brightness, etc., as well as for luminous flux. The brightness of a total radiator would be given in these units by $\frac{1}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$; and the intensity in a given direction would be $\frac{A'}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$, where A' is the projected area of the total radiator on a plane perpendicular to the given direction. However, such physically-defined units are not in use. In a previous paper (Caldin, 1945), the values of $B = \frac{1}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$ for various temperatures were calculated, and given as values of the brightness of a total radiator expressed in lightwatts per cm.^{2} of projected area (more accurately, lightwatts cm.$^{-2}$ steradian^{-1}). But this is not a recognized practice among photometrists, and in order not to raise here the whole question of physical definition of photometric units we shall use the international candle as the unit of intensity, and the lumen as the unit of luminous flux (cf. *International Lighting Vocabulary*, C.I.E. 1938). The ratio of the lumen to the lightwatt, M, will therefore appear in the equations. Thus, the total luminous flux from unit area of a total radiator is $M\int_0^\infty V_\lambda J_\lambda d\lambda$ lumens; the brightness B_c is $\frac{M}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$ candles per cm.^{2} of projected area; and the intensity in a given direction is $\frac{A'M}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$ candles, where A' is the projected area of the source on a plane perpendicular to that direction. The ratio M is somewhat uncertain, but we shall not need to insert a numerical value for it for most of the purposes of this paper. Values of M for various values of the radiation constants, and a general method of computation, have been given in a previous paper (Caldin, 1946).

The term ***Planckian radiator*** is used in this paper to denote a radiator whose emissivity is constant with respect to wave-length and temperature, though not necessarily equal to unity. The energy flux from such a body at any wave-length and temperature bears a constant ratio to that from a total radiator at the same wave-length and temperature. The term thus covers both "black" and "grey" bodies.

§2. DERIVATION OF EXPRESSIONS FOR LIGHT EMISSION DURING COOLING OF A PLANCKIAN RADIATOR

We consider a Planckian radiator, cooling by radiation alone. Its temperature will be assumed uniform. The emissivity of the body will be assumed to be independent of temperature and wave-length, and the specific heat independent of temperature, at the relevant temperatures. The surface is assumed to be a perfect diffuser.

Let

m = mass of body (gm.)

A = surface area of body (cm.^{2}).

s = specific heat of body (cals. gm.$^{-1}$ deg. K^{-1}).

E = emissivity.

t = time from start of cooling, in seconds.

H = total heat content of body, at time t sec. (cals.).

T = temperature of body at time t sec. (° K.).

T_0 = initial temperature of body at $t=0$.

T_s = temperature of surroundings, assumed constant, and $<T_0$.

$B=\frac{1}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$, where $J_\lambda d\lambda$ is in watts cm.$^{-2}$

B_e = brightness of body = luminous intensity per unit projected surface area, in a given direction, in candles cm.$^{-2}=MB=\frac{M}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda$.

M = ratio of lumen to lightwatt.

I = luminous intensity, in a given direction (candles).

$W=\int_0^\infty B_e dt$ = integral with respect to time of the intensity per unit projected area in a given direction; the integral being taken over the whole duration of the cooling (candle-sec.).

$\boldsymbol{K}$ = Stefan's constant.

$k=\boldsymbol{K}AE/ms$.

By Stefan's law we have, then, for the net rate of energy loss from the body by radiation,

$$-dH/dt=\boldsymbol{K}AE(T^4-T_s^4). \qquad \ldots\ldots(1)$$

We have, also, since the temperature of the body is assumed to remain uniform,

$$dH=ms\,dt. \qquad \ldots\ldots(2)$$

Substituting for dH in equation (1),

$$-dT/dt=\frac{\boldsymbol{K}AE}{ms}(T^4-T_s^4),$$

or, writing

$$k = KAE/ms,$$

$$-dT/dt = k(T^4 - T_s^4). \qquad \ldots\ldots(3)$$

This may be integrated by partial fractions, giving

$$t = \frac{1}{4kT_s^3}\left[12\tan^{-1} T/T_s + \ln(T+T_s)/(T-T_s)\right]_{T_0}^{T}. \qquad \ldots\ldots(4)$$

On expanding the two terms in the bracket as series, and inserting the limits T_0 and T, we obtain the following relation:

$$t = [(T^{-3} - T_0^{-3})/3k] + [T_s^4(T^{-7} - T_0^{-7})/7k] + \ldots \qquad \ldots\ldots(5)$$

All the terms of this equation after the first may generally be neglected when T_s is of the order of room temperature; thus, for example, when $T_0 = 2640°$ K., $T = 2000°$ K., and $T_s = 300°$ K., the second term is only 1/3000 of the first, and later terms will be still smaller. Thus, as a very good approximation, we may write

$$T = (3kt + T_0^{-3})^{-\frac{1}{3}}. \qquad \ldots\ldots(6)$$

From this relation between the temperature and time, a relation between the brightness and time is found as follows. In a previous paper (Caldin, 1945) values of B $\left(\text{i.e. } \frac{1}{\pi}\int_0^\infty V_\lambda J_\lambda d\lambda\right)$ were given for a total radiator at a series of temperatures from 2000° K. to 3120° K.; these were calculated by combining Planck's energy-wave-length-temperature relation with the C.I.E. standard luminosity function. It was found that B could be expressed as a function of temperature (T) by an equation of the form

$$\log_{10} B = \alpha - \beta/T, \qquad \ldots\ldots(7)$$

where α and β are constants which depend on the units used for B and T and on the values of the constants C_1 and C_2. Using Birge's 1941 values for C_1 and C_2 ($C_1 = 3{\cdot}7403 \times 10^{-5}$ erg. sec.$^{-1}$ cm.$^{+2}$, $C_2 = 14384{\cdot}_8$ micron degrees), and expressing T in degrees K., the values of the constants are: $\alpha = 4{\cdot}279$, $\beta = 10902$. (With the values $C_1 = 3{\cdot}703 \times 10^{-5}$, $C_2 = 14330$ (*International Critical Tables*), we find $\alpha = 4{\cdot}275$, $\beta = 10860$.)

For a source of emissivity E (assumed constant with respect to temperature and wave-length), the equation corresponding to (7) can be written

$$B = aE\exp(-c/T), \qquad \ldots\ldots(8)$$

where $\log_{10} a = \alpha$, and $c = \beta/0{\cdot}43429$. With Birge's values of C_1 and C_2, $c = 25103$ when T is in ° K., and $a = 1{\cdot}90_1 \times 10^4$. These values will be referred to as c_1 and a_1.

Hence for the brightness, expressed in candles per cm.^{2}, we obtain

$$B_c = MB = aEM\exp(-c/T). \qquad \ldots\ldots(9)$$

A relation between the brightness of a cooling total radiator initially at temperature T_0, and the time t which has elapsed since cooling began, is obtained by combining equations (6) and (9). The result is

$$B_c = aEM\exp[-c(3kt + T_0^{-3})^{\frac{1}{3}}]. \qquad \ldots\ldots(10)$$

The definite integral of this expression with respect to time from $t=0$ to $t=\infty$, corresponding to the temperature limits T_0 to T_s, is:

$$W=\int_0^\infty B_c\,dt=\frac{aEM}{kc^3}\,[(u^2+2u+2)\exp(-u)], \qquad \ldots\ldots(11)$$

where $u=c/T_0$. This integration presupposes that the relation (9) holds with sufficient accuracy down to T_s, which may be well below 2000° K., the lower limit for the calculated values of B on which equation (9) is based. We shall later estimate the accuracy and show that it is in fact sufficient when T_0 is of the order of 3000° K.

On replacing k by $\boldsymbol{K}AE/ms$, and writing $f(u)$ for $[(u^2+2u+2)\exp(-u)]$, and G for $a/c^3\boldsymbol{K}$., equation (11) becomes

$$W=MG(ms/A)f(u). \qquad \ldots\ldots(12)$$

If we insert the values given above for a and c, i.e. a_1 and c_1, corresponding to Birge's 1941 values of C_1 and C_2, and his value of $\boldsymbol{K}$, namely $1{\cdot}354\times10^{-12}$ cal. cm.$^{-2}$ deg.$^{-4}$ sec.$^{-1}$, we find $G=887$ (dimensions: °K.). This value will be referred to as G_1. (With $C_1=3{\cdot}703\times10^{-5}$ and $C_2=14330$, we find $G=889$.)

Equations (9) and (12) both refer to the luminous flux in a given direction due to unit projected area of the radiating source. To obtain the luminous intensity (in candles) in a given direction due to the whole body, at a given time, and the integral with respect to time of the intensity (in candle-sec.), we must multiply B_c and W respectively by A', the projected area of the body on a plane perpendicular to the given direction.

Thus finally, for the ***intensity at time t*** we obtain

$$I=A'B_c=aEA'M\exp[-c(3kt+T_0^{-3})^{\frac{1}{3}}], \qquad \ldots\ldots(13)$$

where I = luminous intensity in candles and the other symbols have already bee defined.

From (13) the *initial* (*maximum*) *intensity* is seen to be

$$I_{\max}=aEA'M\exp(-c/T_0)=aEA'M\exp(-u). \qquad \ldots\ldots(14)$$

and the ***intensity at time t, relative to the maximum***, is

$$I/I_{\max}=\exp[c/T_0-c(3kt+T_0^{-3})^{\frac{1}{3}}]. \qquad \ldots\ldots(15)$$

For the *time-integral of the intensity* we obtain

$$\int_0^\infty I\,dt=A'\int_0^\infty B_c\,dt=A'W=MG\,.\,ms(A'/A)f(u). \qquad \ldots\ldots(16)$$

Thus the integral of the intensity in a given direction with respect to time ($A'W$)—which is the main factor controlling the total actinic effect of such a source as we are considering—depends, according to equation (16), on the initial temperature (since $c/u=T_0$), the heat capacity (ms), and the ratio A'/A, which depends on the shape of the body (for a sphere it is 1/4). It does not depend on the emissivity. The initial temperature has the greatest proportional influence; whereas the value of $A'W$ is directly proportional to the heat capacity and to A'/A, a small change of initial temperature in the region of 2500° K. (for example) produces a change in $A'W$ proportionally about eight times greater.

The initial intensity in a given direction, $I_{\max}$, depends, according to equation (14), on the emissivity (E), projected area (A'), and initial temperature. The shape of the intensity-time curve depends, according to equation (15), on the emissivity, total surface area, heat capacity (since $k = \boldsymbol{K}AE/ms$), and initial temperature; the time required for the intensity to fall to a given fraction of its initial value, for a given value of T_0, is proportional to $1/k$, and, therefore, increases with (*a*) decrease of emissivity, E; (*b*) decrease of specific surface area, A/m; and (*c*) increase of specific heat, s.

If we are considering the light emitted by a body whose mass, dimensions, and specific heat are known, but whose initial temperature and emissivity are unknown—as generally when we are dealing with the experimental data on these transient sources—we can proceed as follows. The observed value of the intensity-time integral, with the mass, specific heat and shape of the body, gives by equation (16) the value of $f(u)$, and hence of the initial temperature T_0. The observed initial intensity then gives the emissivity E, by equation (18). Then the whole intensity-time curve, or relative intensity-time curve, can be calculated from equation (13) or (15), and compared with the observed curve. For a source of given heat capacity and dimensions, a choice of values for k and T_0 is enough to fix the whole theoretical intensity-time curve (and its integral), since the value of E corresponding to the chosen values of k is at once calculable from the relation $k = \boldsymbol{K}AE/ms$, and all the required constants are then known.

§ 3. NUMERICAL VALUES FOR TIME-INTENSITY INTEGRAL

In the table below are given the calculated values of $f(u) = f(c/T_0)$ at a series of temperatures from 2000° K. to 6000° K., with the value c_1 for c. Numerical values of the time-intensity integral may be obtained by inserting these values in equation (16). In the table are given also values of $671G_1 f(u)/4$, which, as may

Table 1

T (°K.)	$f(c_1/T_0)$	$(671G_1 f(c_1/T_0)/4) \times 10^3$
2000	0·000654	0·097
2080	0·000984	0·146
2160	0·00144	0·214
2240	0·00204	0·303
2320	0·00281	0·418
2400	0·00379	0·564
2480	0·00501	0·745
2560	0·00650	0·966
2640	0·00826	1·23
2720	0·0104	1·54
2800	0·0128	1·91
2880	0·0156	2·33
2960	0·0188	2·80
3040	0·0225	3·34
3120	0·0265	3·95
4000	0·0977	14·5
5000	0·233	34·6
6000	0·394	58·6

be seen from equation (16), is the value obtained for the time-intensity integral, in candle-seconds, for light emitted in any direction from a spherical total radiator of unit heat capacity, cooling from a given temperature T_0, if we adopt the value G_1 for G and the corresponding value 671 for M (Caldin, 1946).

§ 4. ACCURACY OF FORMULAE (13) TO (16)

Accuracy when T_0 is less than 3120° K.

It remains to consider the accuracy of the values of I and of $\int_0^\infty I\,dt$ calculated from equations (13) to (16). The percentage error in I is equal to the percentage error in the value of B calculated from equation (8), with the values of a and c assumed above; this, as already stated, is not more than 1%. The error in W is made up of an error equal to this and an error due to the assumption that the relation (8), which has been tested by calculation only down to 2000° K., holds also between 2000° K. and the final temperature T_s, with the same values of a and c,

Actually the relation (8), with the constants a_1 and c_1 given above, begins to deviate appreciably from the truth below 2000° K. This may be shown by comparing values of B calculated from it with the values calculated by Ives (1926). Ives used the C.I.E. relative luminosity data and the value 14330 micron degrees for C_2; however, instead of using the actual C.I.E. values, he used a formula which reproduced them with an error of the order of 1%; his values of B are subject to the same error. He gives values of $B_c = \frac{M}{\pi}\int_0^\infty V_\lambda J_\lambda\,d\lambda$ with the value $M = 1/0{\cdot}00161$. For comparison with the results of equation (8), the corresponding values of $B = B_c/M$ have been computed. These values are compared in the following table with those calculated from the equation

Table 2

T (°K.)	$\log_{10} B$ (Ives) (*a*)	$\log_{10} B$ from eqn. (17) (*b*)	Difference (*a*)−(*b*)	$\log_{10} B$ from eqn. (22) (*c*)	Difference (*a*)−(*c*)
2000	$\bar{2}{\cdot}854$	$\bar{2}{\cdot}845$	+0·009	$\bar{2}{\cdot}853$	+0·001
1900	$\bar{2}{\cdot}572$	$\bar{2}{\cdot}559$	+0·013	$\bar{2}{\cdot}573$	−0·001
1800	$\bar{2}{\cdot}260$	$\bar{2}{\cdot}242$	+0·018	$\bar{2}{\cdot}262$	−0·002
1700	$\bar{3}{\cdot}914$	$\bar{3}{\cdot}887$	+0·027	$\bar{3}{\cdot}915$	−0·001
1600	$\bar{3}{\cdot}525$	$\bar{3}{\cdot}487$	+0·038	$\bar{3}{\cdot}524$	+0·001
1400	$\bar{4}{\cdot}591$	$\bar{4}{\cdot}518$	+0·073	$\bar{4}{\cdot}575$	+0·016
1200	$\bar{5}{\cdot}356$	$\bar{5}{\cdot}225$	+0·131	$\bar{5}{\cdot}310$	+0·046

$$\log_{10} B = 4{\cdot}275 - 10860/T, \qquad \text{......(17)}$$

in which the constant 10860 corresponds with $C_2 = 14330$ micron degrees. The fourth column gives the difference in log units.

It is clear that the divergence between the directly calculated values and those derived from equation (17) increases with decrease of temperature; it becomes as much as 9% at 1600° K. Since, however, very little light is emitted at such low temperatures, the error in the time-intensity integral will be much less than this. This error is best estimated as follows. Ives' values from 2000° K. down to 1600° K. are represented within 0·2% by the equation

$$\log_{10} B = 4{\cdot}168 - 10630/T, \qquad \ldots\ldots(18)$$

as may be seen from the sixth column of the table above. We can thus use equation (18) below 2000° K., down to the lowest temperature which need be considered, with the same accuracy as that afforded by equation (17) between 2000° K. and 3120° K. Correcting equation (18) to correspond with Birge's 1941 values of C_1 and C_2, we get for temperatures below 2000° K.

$$\log_{10} B = 4{\cdot}172 - 10670/T, \qquad \ldots\ldots(19)$$

and the constants in equation (8) become $a = 1{\cdot}49 \times 10^4$, $c = 24568$. This gives G in equation (16) as 740. Let these values of the constants be designated a_2, c_2, G_2, and the values found for the temperature range 2000° to 3120° K. be a_1, c_1, G_1. Then the value of the time-intensity integral for light emitted below 2000° K. is, by equation (16), $[G_2 ms(A'/A) f(c_2/2000)]$. The corresponding value obtained by assuming the values of the constants a_1, c_1 and G_1 to hold over the whole temperature range T_0 to T_s is $[G_1 ms(A'/A) f(c_1/2000)]$. Thus the error in using a_1, c_1 and G_1 below 2000° K. would be

$$[ms\,.\,A'/A][G_1 f(c_1/2000) - G_2 f(c_2/2000)] \qquad \ldots\ldots(20)$$

and the percentage error in the whole value of the time-intensity integral for a body cooling from a temperature T_0 would be

$$100\,[f(c_1/2000) - \frac{G_2}{G_1} f(c_2/2000)]/f(c_1/T_0). \qquad \ldots\ldots(21)$$

Inserting numerical values, this gives

$$\text{percentage error} = 0{\cdot}003/f(c_1 T_0). \qquad \ldots\ldots(22)$$

Thus the percentage error depends only on T_0, not on k; and, as would be expected, it decreases with increase of T_0. For $T_0 = ca.$ 3000° K., $f(c_1/T_0)$ has the value *ca.* 0·02, so that the time-intensity integral for light emitted below 2000° K. is about 3% of the total (this fraction depends only on T_0), and the error due to the use of the values a_1 and c_1 throughout is 0·15%. The total error in W would then be less than 1%.

Accuracy with higher values of T_0.

So far, only values of T_0 lower than 3120° K. have been considered. To estimate the additional errors in using formula (13) to (16) with higher values of T_0 and the same values a_1, c_1 and G_1 for the constants, values of B calculated from equation (17) may be compared with those derived from Ives' calculated values of B_c for higher temperatures, as in table 3.

Thus the error in calculating B from equation (17) is under 1% with $T_0 = 3000°$ K.; 3% with $T_0 = 4000°$; 7% with $T_0 = 5000°$, and 9% with $T_0 = 6000°$ K. The percentage errors in calculating I from equation (13), or $I_{\max}$ from equation (14), will correspond with these errors. The percentage error in the time-intensity integral will be less; we can estimate it by a method somewhat similar to that of the previous section. Values of B derived from Ives' values of B_c from 3000° K. to 6000° K. can be represented within about $\pm 1\,\%$ by

$$\log_{10} B = 4{\cdot}339 - 11040/T \qquad \ldots\ldots(23)$$

in place of equation (17) or (18). Converting the constants to correspond with Birge's 1941 values of C_1 and C_2, we find that the corresponding values of the constants a, c and G are $a_3 = 2{\cdot}20_3 \times 10^4$, $c_3 = 25518$, $G_3 = 980$. It may then be easily shown that the additional percentage error due to using a_1, c_1, and G_1 in calculating the time-intensity integral is

$$\frac{100[G_1 f(c_1/T_0) - G_3 f(c_3/T_0) - G_1 f(c_1/3000) + G_3 f(c_3/3000)]}{[G_1 f(c_1/3000) - G_3 f(c_3/3000) + G_3 f(c_3/T_0)]}. \qquad \ldots\ldots(24)$$

Inserting numerical values, we find, for $T_0 = 4000$° K., $-2{\cdot}6\,\%$; for 5000° K., $-4{\cdot}3\,\%$; for 6000° K., $-5{\cdot}4\,\%$.

Table 3

T (°K.)	$\log_{10} B$ (Ives) (*a*)	$\log_{10} B$ from eqn. (17) (*b*)	Difference (*a*)−(*b*)
3000	0·659	0·655	+0·004
4000	1·574	1·560	+0·014
5000	2·131	2·103	+0·028
6000	2·504	2·465	+0·039

ACKNOWLEDGMENT

This paper is based on work carried out in the Ministry of Supply. The author is indebted to Mr. J. S. Dick for facilities and encouragement, to the Chief Scientific Officer, Ministry of Supply, for permission to publish, and to Mr. J. S. Preston and Dr. R. F. Barrow for their suggestions.

REFERENCES

Beck and Eggert, 1926. *Z. wiss. Photogr.*, **24**, 376.

Birge, R. T., 1941. *Rev. Mod. Phys.*, **13**, 233.

Caldin, E. F., 1945. *Proc. Phys. Soc.*, **57**, 440 ; 1946. *Proc. Phys. Soc.*, **58**, 207.

Eder, J. M., 1930. *Z. wiss. Photogr.*, **27**, 337.

Hyde, Forsythe and Cady, 1919. *Phys. Rev.*, **13**, 45.

Ives, H. E., 1926. *J. Opt. Soc. Amer.*, **12**, 75.

Van Liempt and de Vriend, 1934. *Rec. Trav. Chim. Pays-Bas*, **53**, 839 ; 1935. *Ibid.*, **54**, 239 ; 1937 : *Physica*, **4**, 353.

Kelley, 1938. *J. Opt. Soc. Amer.*, **28**, 27.

Moon, P., 1936. ***Scientific Basis of Illuminating Engineering*** (McGraw-Hill), pp. 53, 57, 552.

Projector, T. H. and Barbrow, L. E., 1945. *Rev. Sci. Instrum.*, **16**, 51.

Reeb, 1935. *Z. wiss. Photogr.*, **34**, 77.

THE DETERMINATION OF THE INITIAL TEMPERATURE OF A COOLING TOTAL RADIATOR FROM MEASUREMENTS OF THE SPECTRAL DISTRIBUTION OF THE ENERGY EMITTED DURING COOLING

By E. F. CALDIN,

Department of Physical and Inorganic Chemistry, Leeds University

MS. received 7 *December* 1945 ; *read* 15 *March* 1946

ABSTRACT. An expression is derived theoretically relating the initial temperature (T_0) of a total radiator, cooling by radiation, to the spectral distribution of the total energy emitted during cooling (i.e. the variation with wave-length of the integral with respect to time of the energy flux at a given wave-length). This expression allows the determination of the initial temperature of such a cooling source (to which photographic flash bulbs, etc., provide approximations), from the spectral distribution of the energy received by a photographic plate or other recording device exposed to the source. This spectral distribution is compared with that of the energy flux from a total radiator at various temperatures ; it is found that for each value of T_0 (between 2500° K. and 4500° K.) a temperature (T_1) can be found such that there is close agreement between the two spectral distributions. T_1 corresponds to the " temperature " cited in the literature for such sources. A linear relation is found to hold between T_1 and T_0.

§ 1. INTRODUCTION

In a previous paper (Caldin, 1946 b), expressions were derived for the time-intensity relation, and the corresponding time-intensity integral, for light emission from a total radiator of uniform temperature, cooling by radiation. The calculations are of interest because the results can be compared with those found experimentally for certain light sources which give continuous spectra and may owe their emission to cooling solid incandescent particles. Examples include certain sources of short duration, such as those produced by combustion of metal wires or foil in oxygen (as in photographic flash bulbs), or by combustion of photographic flash powders; and also some continuous sources, e.g. the " Nitralamp ", a projection lamp in which magnesium ribbon is burned. Besides parameters such as the mass, shape and emissivity of the source, the expressions derived contain T_0, the initial temperature. This is not readily accessible, though in principle it might be obtained by continuously recording, by means of cathode-ray oscillographs, the relative intensities of the light in, say, the red and blue regions of the spectrum. Values given for the temperatures of these sources have usually been obtained by burning the flash powder, etc., in front of the slit of a spectrograph, exposing the plate to the flash during the whole of its duration, and comparing the distribution of density along the spectrogram with the density-distributions produced by sources emitting approximately Planckian radiation at various constant temperatures. It is evident that, since the temperature is falling continuously, the spectral distribution of the total energy which falls on the plate (i.e. the spectral distribution of the integral of the energy flux with respect to time) will not be the same as the spectral distribution corresponding to the initial temperature of the source, and the derived "temperature" will be lower than T_0. The following values for the "temperatures" of such sources have been recorded:—

Source.	"Temperature" (° K.)	Reference.
Aluminium foil in oxygen (photoflash lamp)	3500	Forsythe and Easley (1934)
Magnesium ribbon in air	3700 3750	Dziobek (1928); O'Brien and Russell (1933)
Magnesium ribbon in oxygen	*ca.* 4000	Eder (1930)
Thorium nitrate-magnesium mixture (equal proportions) = Agfa flash powder.	3000–3100 3350	Eder (1930); Arens and Eggert (1930)

It has not, however, been shown that the radiation from such sources is strictly Planckian; indeed, the likelihood of variation of emissivity, and of emission of band spectra from AlO and MgO, make this very doubtful.

In this paper, a relation is sought between the initial temperature T_0 of a total radiator, which cools by radiation alone, and in which temperature equilibrium is maintained, and the "temperature", T_1, derived from the spectral distribution of the time integral of the energy flux. The latter distribution is calculated in terms of T_0, using the time-temperature relation given in the preceding paper. It is found that the distribution is determined by T_0 alone and does not depend on any other of those properties of the source which control its rate of cooling. It does not correspond exactly with the spectral distribution of the energy flux from a Planckian radiator at any single temperature T_1; but a value of T_1 can be found such that the spectral distribution of the energy flux of a total radiator at T_1° K. agrees quite closely with the spectral distribution of the time integral of the energy flux from a cooling radiator initially at some higher temperature T_0° K. over the whole visible range, at any rate when T_0 is below 4000° K. Values of T_1, for various values of T_0, are obtained, and a relation between T_1 and T_0 derived. It is thus possible, from any given value of T_1, to find T_0, the initial temperature; no knowledge of the emissivity, area, specific heat, etc., of the source is required, provided only that the emissivity is constant with respect to temperature and wave-length.

Work is in progress in several laboratories to determine the spectral distribution both of the time-integral of the energy-flux and of the energy-flux at the maximum temperature, for the light emitted by aluminium foil burning in oxygen. The calculations of this paper should be of assistance in interpreting the experimental results.

§ 2. CALCULATION OF SPECTRAL DISTRIBUTION OF TOTAL ENERGY (INTEGRAL OF ENERGY FLUX WITH RESPECT TO TIME) EMITTED DURING COOLING OF A TOTAL RADIATOR

The spectral distribution of the energy-flux from a total radiator at temperature T is given by Planck's equation:

$$J_\lambda = \frac{C_1\lambda^{-5}}{e^{C_2/\lambda T}-1}, \qquad \ldots\ldots(1)$$

where C_1 and C_2 are constants, and $J_\lambda d\lambda$ is the total energy-flux from unit area of a total radiator (watt cm.$^{-2}$) between the wave-lengths $\lambda \pm \frac{d\lambda}{2}$ (microns). The simplified form of this equation, due to Wien, holds for wave-lengths up to $0{\cdot}7\,\mu$ within (for example) about 0·3 % at 3000° K., and within about 2 % at 5000° K:

$$J_\lambda = C_1\lambda^{-5}\exp\left(-C_2/\lambda T\right). \qquad \ldots\ldots(2)$$

The relation between temperature and time, for a total radiator which cools by radiation alone, and in which temperature equilibrium is maintained, is given to a close approximation by the following equation (Caldin, 1946)

$$T = (3kt + T_0^{-3})^{-1/3} \qquad \ldots\ldots(3)$$

where

T = temperature at time t (° K.); T_0 = temperature at zero time;
$k = \boldsymbol{K}AE/ms$, where
$\boldsymbol{K}$ = Stefan's constant;
A = surface area of the radiating body (cm^2);
E = emissivity of body, assumed constant with respect to wave-length and temperature;
m = mass of body (gm.); and
s = specific heat of body, assumed constant with respect to temperature (cal. gm.$^{-1}$ deg. K.$^{-1}$).

Substituting in equation (2) the value of T given by equation (3), we find for the variation with time of the energy flux per cm^2 per micron at a given wave-length

$$J_\lambda = C_1\lambda^{-5}\exp\left[-\frac{C_2}{\lambda}(3kt + T_0^{-3})^{1/3}\right]. \qquad \ldots\ldots(4)$$

The definite integral of this expression with respect to time, from $t=0$ to $t=\infty$, is

$$\int_0^\infty J_\lambda dt = \frac{C_1\lambda^{-4}}{C_2kT_0^2}\left[1 + \frac{2\lambda T_0}{C_2} + 2\left(\frac{\lambda T_0}{C_2}\right)^2\right]\exp\left(-C_2/\lambda T_0\right). \qquad \ldots\ldots(5)$$

This is the required expression relating the wave-length to the integral with respect to time of the corresponding energy-flux per micron per cm^2.

It may be compared with equation (2) and we see that whereas, at a given temperature T_0, the variation J_λ with wave-length is given by the expression

$$J_{\lambda, T_0} \propto \left[\lambda^{-5}\exp\left(-C_2/\lambda T_0\right)\right], \qquad \ldots\ldots(6)$$

the variation with wave-length of the time integral of J_λ, for cooling from T_0, is given by an expression containing an extra term:

$$\int_0^\infty J_\lambda dt \propto \left[\lambda^{-5}\exp(-C_2/\lambda T_0)\right]\left[\lambda\left\{1+\frac{2\lambda T_0}{C_2}+2\left(\frac{\lambda T_0}{C_2}\right)^2\right\}\right]. \qquad \ldots\ldots(7)$$

It is convenient to express the spectral distribution with respect to unity at some wave-length. Thus, taking the values relative to that at $\lambda=\lambda_1=0\cdot59\,\mu$, and combining (6) and (7), we obtain:

$$\int_0^\infty J_\lambda\, dt \Big/ \int_0^\infty J_{0\cdot59}\, dt=[J_{\lambda,\,T_0}/J_{0\cdot59,\,T_0}]\,[\lambda(1+2v+2v^2)/\lambda_1(1+2v_1+2v_1^2)], \qquad \ldots\ldots(8)$$

where $$v=\lambda T_0/C_2, \quad v_1=\lambda_1 T_0/C_2.$$

The values of the second bracket in the right-hand side of equation (8) are a measure of the difference between the spectral distribution of the energy radiated, to a bolometer or a photographic plate, from a total radiator maintained at temperature T_0, and that of the integral with respect to time of the energy flux from the same radiator while cooling from T_0. It will be seen that the latter energy-distribution depends only on the initial temperature, T_0, and not on the parameter k, which is also concerned in determining the rate of cooling. Thus for any value of T_0 we can plot curves of $\left(\int_0^\infty J_\lambda\, dt \Big/ \int_0^\infty J_{0\cdot59}\, dt\right)_{T_0}$ against λ, which will be valid for all total radiators, provided that the assumptions implied in the use of equation (3) are fulfilled. These curves may be compared with the curve of $(J_\lambda/J_{0\cdot59})_T$ against λ, i.e. with the spectral distribution of the energy-flux at a given temperature, T, for various values of T. If the two spectral distributions agree fairly closely for some value of T, say T_1, then T_1 is the "temperature" of the transient source obtained according to the methods mentioned in the introduction.

§ 3. RESULTS AND DISCUSSION

Calculations of $\left(\int_0^\infty J_\lambda\, dt \Big/ \int_0^\infty J_{0\cdot59}\, dt\right)_{T_0}$ have been carried out for the following values of T_0:—2500°, 3000°, 3500°, 4000° and 4500° K. The spectral-energy distribution tables of Skogland (1929) and Frehafer and Snow (1926) were used as sources of values of $J_\lambda/J_{0\cdot59}$ at a series of wave-lengths in the visible region Skogland's tables, using the value $C_2=$ 14330 micron degrees, cover the range 2000° K. to 3120° K. in steps of 20°; they have been used for temperatures up to 3120° K., with the corresponding value of C_2. Frehafer and Snow's tables, using the value $C_2=14350$ micron degrees, cover the range 1000° to 28000° K.; they have been used, with the corresponding value of C_2, for temperatures above 3120° K.; their interval is 250° in the relevant range. To make the two sets of data accurately comparable, all the temperatures used in connection with Skogland's tables might be multiplied by 14350/14320, but the corrections are at most six or seven degrees, and are smaller than the uncertainty in T_1; they have therefore been neglected.

The data obtained for the curves of $\left(\int_0^\infty J_\lambda dt \Big/ \int_0^\infty J_{0\cdot 59}\, dt\right)_{T_0}$ for the several values of T_0 are given in table 1 below. For each value of T_0, there is given for comparison the data for curves of $(J_\lambda/J_{0\cdot 59})_T$ against λ for suitable values of T, in the region for which the two spectral distributions are similar. Except for $T_0=4500°$ K., the middle value given for T is T_1, the value of T which gives the closest agreement between the spectral distribution of the time-integral of the energy-flux of a source cooling from T_0, and that of the energy-flux of a source

Table 1

(i)

$T_0=2500°$ K.

λ (μ)	$\dfrac{\int_0^\infty J_\lambda\, dt}{\int_0^\infty J_{0\cdot 59}\, dt}$	$\dfrac{J_\lambda}{J_{0\cdot 59}}$		
	$T_0=$ 2500° K.	$T=$ 2180	$T=$ 2200	$T=$ 2220
0·32	0·0029	0·0018	0·0019	0·0021
0·40	0·044	0·035	0·037	0·039
0·45	0·137	0·121	0·125	0·129
0·50	0·327	0·308	0·314	0·319
0·55	0·644	0·632	0·636	0·641
0·59	1·000	1·000	1·000	1·000
0·65	1·699	1·723	1·707	1·692
0·70	2·413	2·450	2·411	2·374
0·72	2·726	2·762	2·712	2·664
0·76	3·383	3·408	3·332	3·258

(ii)

$T_0=3000°$ K.

λ (μ)	$\dfrac{\int_0^\infty J_\lambda\, dt}{\int_0^\infty J_{0\cdot 59}\, dt}$	$\dfrac{J_\lambda}{J_{0\cdot 59}}$		
	T_0 3000° K.	$T=$ 2540	$T=$ 2560	$T=$ 2580
0·32	0·011	0·0067	0·0071	0·0076
0·40	0·094	0·074	0·077	0·080
0·45	0·225	0·198	0·202	0·207
0·50	0·435	0·409	0·415	0·420
0·55	0·723	0·709	0·712	0·716
0·59	1·000	1·000	1·000	1·000
0·65	1·470	1·490	1·479	1·469
0·70	1·884	1·912	1·890	1·868
0·72	2·053	2·078	2·050	2·023
0·76	2·384	2·395	2·356	2·317

(iii)

$T_0=3500°$ K.

λ (μ)	$\dfrac{\int_0^\infty J_\lambda\, dt}{\int_0^\infty J_{0\cdot 59}\, dt}$	$\dfrac{J_\lambda}{J_{0\cdot 59}}$		
	$T_0=$ 3500° K.	$T=$ 2840	$T=$ 2900	$T=$ 2960
0·32	—	0·016	0·018	0·021
0·40	0·122	0·120	0·134	0·142
0·45	0·318	0·271	0·286	0·302
0·50	0·531	0·491	0·507	0·522
0·55	0·783	0·763	0·773	0·782
0·59	1·000	1·000	1·000	1·000
0·65	1·323	1·357	1·335	1·315
0·70	1·586	1·632	1·587	1·546
0·72	1·684	1·732	1·678	1·627
0·76	—	1·912	1·832	1·770

(iv)

$T_0=4000°$ K.

λ (μ)	$\dfrac{\int_0^\infty J_\lambda\, dt}{\int_0^\infty J_{0\cdot 59}\, dt}$	$\dfrac{J}{J_{0\cdot 59}}$		
	$T_0=$ 4000° K.	$T=$ 3120	$T=$ 3250	$T=$ 3500
0·32	—	—	—	—
0·40	0·190	0·173	0·200	0·257
0·45	0·412	0·344	0·378	0·446
0·50	0·615	0·563	0·595	0·654
0·55	0·831	0·806	0·824	0·856
0·59	1·000	1·000	1·000	1·000
0·65	1·230	1·265	1·230	1·171
0·70	1·394	1·447	1·381	1·270
0·72	1·454	1·509	1·429	1·298
0·76	—	—	—	—

(v)

$T_0=4500°$ K.				
λ (μ)	$\dfrac{\int_0^\infty J_\lambda\,dt}{\int_0^\infty J_{0\cdot59}\,dt}$	$\dfrac{J_\lambda}{J_{0\cdot59}}$		
	$T_0=$ 4500° K.	$T=$ 3250	$T=$ 3500	$T=$ 3750
0·32	—	—	—	—
0·40	0·245	0·200	0·257	0·320
0·45	0·504	0·378	0·446	0·514
0·50	0·692	0·595	0·654	0·712
0·55	0·871	0·824	0·856	0·886
0·59	1·000	1·000	1·000	1·000
0·65	1·161	1·230	1·171	1·123
0·70	1·266	1·381	1·270	1·182
0·72	1·302	1·429	1·298	1·196
0·76	—	—	—	—

at constant temperature T, over the visible range. The other two are values, on either side of T_1, for which the agreement definitely appears inferior; they give some measure of the uncertainty in T_1.

It will be seen that, for each temperature T_0, some temperature T_1 can be found such that the two spectral distributions agree fairly closely. The discrepancies (with the best values of T_1) increase with T_0, and so does the uncertainty in T_1. Thus, when $T_0=2500°$ K. the maximum discrepancy (over the wave-length range considered) is about 1 % of the value at $0\cdot59\,\mu$, and T_1 can be determined within $\pm 20°$; but when $T_0=4500°$ K., the maximum discrepancy is about 4 % of the value at $0\cdot59\,\mu$, and T_1 can only be fixed within about $\pm 150°$ (agreement at the lower wave-lengths is best with $T_1=3750°$ K., at the higher wave-lengths with $T_1=3500°$ K.). It is, however, evident that, at least up to $T_0=4000°$ K. the spectral distribution of the time-integral of the energy flux will agree within the usual experimental error with the spectral distribution of the energy-flux for some temperature T_1. The values of T_1 corresponding to various values of T_0, and the approximate limits of the uncertainties in T_1, are summarized in the following table:—

Table 2

T_0 (° K.)	T_1 (° K.)	T_1 calc. from (9) ($450+0\cdot70\ T_0$)
2500	2200 ± 20	2200
3000	2560 ± 20	2550
3500	2900 ± 50	2900
4000	3250 ± 100	3250
4500	3600 ± 150	3600

These values of T_0 and T_1 obey closely the following linear relation:

$$T_1 = 450 + 0{\cdot}70 T_0, \qquad \ldots\ldots(9)$$

as may be seen from the third column in table 2. Over the range $T_0 = 2500°$ K. to 4500° K., therefore, the value of T_0 can be determined from the value of T_1, corresponding to the experimentally-determined spectral distribution of the time-integral of the energy-flux, by using equation (9).

If the radiation from the sources mentioned in the Introduction is in fact Planckian within small limits, and if the other assumptions of the treatment are justified, the values cited for the "temperature" should be increased by amounts between 500° K. and 1000° K. to obtain the value of the maximum initial temprature T_0. The temperature so obtained for the photoflash lamp is above the boiling point of aluminium oxide, which is in the region of 3000° K. (Ruff and Konschak, 1926). This suggests that the radiation is possibly not Planckian, and that the emissivity decreases with increase of wave-length in the visible range. This would be compatible with the work of Liebmann (1930) on the emissivity of aluminium oxide. Current experimental work on emissivities may throw further light on the matter. Further application of the theory must await the completion of this work and of various studies on the burning of aluminium in oxygen.

REFERENCES

Arens, H. and Eggert, J., 1930. *Z. wiss. Photogr.*, **28**, 169.
Caldin, E. F., 1946 a. *Proc. Phys. Soc.*, **58**, 207.
Caldin, E. F., 1946 b. *Proc. Phys. Soc.*, **58**, 341.
Dziobek, Z. F., 1928. *Z. wiss. Photogr.*, **25**, 287.
Eder, J. M., 1930. *Z. wiss. Photogr.*, **27**, 337.
Forsythe and Easley, 1934. *Phys. Rev.*, **45**, 123.
Frehafer and Snow, 1926. *U.S. Bur. Misc. Pub.* No. 56.
Liebmann, G., 1930. *Z. Phys.*, **63**, 404.
O'Brien and Russell, 1933. *J. Opt. Soc. Amer.*, **23**, 118.
Ruff and Konschak, 1926. *Z. Elektr.*, **32**, 515.
Skogland, J. F., 1929. *U.S. Bur. St. Misc. Pub.* No. 86.

DISCUSSION

on papers by E. F. Caldin entitled:—

"The relation between the brightness and temperature of a total radiator." *Proc. Phys. Soc.*, **57**, 440 (1945).
"The ratio of the new international lumen to the lightwatt." *Ibid.*, **58**, 207 (1946).
"Light emission during cooling of a Planckian radiator." *Ibid.*, **58**, 341 (1946).
"The determination of the initial temperature of a cooling total radiator from measurements of the spectral distribution of the energy emitted during cooling." *Ibid.*, **58**, 350 (1946).

Dr. A. G. Gaydon. I should like to enquire if Mr. Caldin has studied the spectrum of his source in detail. Burning magnesium ribbon emits strong MgO bands, and I should have expected emission of the green system of AlO in the burning of aluminium foil in oxygen. This might affect Mr. Caldin's determinations of the temperature.

Mr. L. T. Minchin. I should like to ask two questions: (*a*) What is a lightwatt? (*b*) Why is the possibility of chemiluminescence excluded from Mr. Caldin's discussion of the Mg flash?

Even if the spectrum is continuous, might not the chemiluminescence spectrum be superimposed on incandescence? Since the body is cooling rapidly whilst the spectral distribution of energy is being recorded, it must be difficult to be certain that it conforms entirely with temperature radiation.

Dr. L. HARTSHORN. I fail to see any excuse for the term " lightwatt ", and hope that physicists will refuse to adopt it. Although the author has disclaimed responsibility for its introduction, by using it he is helping to burden physics with an unnecessary unit. I gather that the quantity concerned is power weighted according to its effect on the eye. However it is weighted, it remains power, and it is difficult to see why the universally accepted unit of power, the watt, is not sufficient. Perhaps some qualifying term like " luminous power " is needed for the quantity, but it is surely out of place in the unit.

Prof. A. O. RANKINE. I agree with others that it is a mistake to use the term " lightwatt " as an apparently new unit, although there is no difference dimensionally from the general unit of power. We do not, and, I hope, will not, speak of " electrical watts ", " mechanical watts ", or " thermal watts ", for which there would also be no justification.

AUTHOR'S reply. (a) ***Mechanism of light emission.*** The purpose of the third and fourth papers under discussion is to derive theoretically the intensity-time curves and spectral distributions of energy to be expected from cooling Planckian radiators. Until experimental investigations now in progress in various laboratories are completed, one cannot adequately compare this theoretical model with experiment. Therefore I do not wish to rule out chemiluminescence as a possible mechanism. The spectroscopic studies of O'Brien and Russell (1933), Forsyth and Easley (1934), and van Liempt and de Vriend (1934, 1935) did not reveal any important contribution to the total energy from MgO or AlO bands.

(b) *Photometric units.* The term " lightwatt " was used by Ives in 1917, and given wider currency by Moon in 1936. If we consider an area A cm.^{2} traversed by a uniform radiant flux, the total energy flux is given by $A\int_0^\infty J_\lambda d\lambda$, where $J_\lambda d\lambda$ is the radiant flux per cm^2 in the wave-length range $\lambda \pm \frac{d\lambda}{2}$. The luminous flux, according to the international definition (*International Lighting Vocabulary, C.I.E.*, 1938) is then $A\int_0^\infty V_\lambda J_\lambda d\lambda$, where V_λ is the value of the relative luminosity factor for the human eye at wave-length λ. (Representative values for V_λ have been internationally accepted since 1924.) If $J_\lambda d\lambda$ is expressed in watts cm^{-2}, the above expression is said to give the luminous flux in lightwatts. The term is thus not analogous to the term " mechanical watts ", to which objection is rightly taken ; for the ratio of $\int_0^\infty V_\lambda J_\lambda d\lambda$ to $\int_0^\infty J_\lambda d\lambda$ depends on the spectral distribution of J_λ, and so is not constant like the ratios of mechanical, electrical, and thermal units of power. The main objection to the use of the term is that the values of V_λ have no special physical status ; they are related to the response of a particular type of receiver under particular conditions, and would be quite different if the receiver were, say, a photoelectric cell. The best specification is, no doubt, a curve or function expressing the spectral distribution of energy flux. But a special term for a frequently occurring integral, closely related to measurable quantities, is always a convenience, and in its absence one has to employ some clumsy device such as that used in the third of the papers under discussion, where intensities computed in terms of $\int_0^\infty V_\lambda J_\lambda d\lambda$ were all multiplied by a constant and expressed in candles. Such computations are of increased importance since the new international candle, defined in terms of the brightness of a total radiator at a fixed temperature, is expected to come into use fairly soon. Perhaps photometrists could suggest a more acceptable term.

REFERENCES

FORSYTH and EASLEY, 1934. *Phys. Rev.*, **45**, 123.
IVES, H. E., 1917. *Astrophys. J.*, **45**, 41.
MOON, 1936. *Scientific Basis of Illuminating Engineering* (McGraw-Hill).
O'BRIEN and RUSSELL, 1933. *J. Opt. Soc. Amer.*, **23**, 118.
VAN LIEMPT and DE VRIEND, 1934. *Rec. Trav. Chim.*, **53**, 839 ; 1935. *Ibid.*, **54**, 239.

THE PROPAGATION OF SOUND BETWEEN WALLS OF POROUS MATERIAL

By R. A. SCOTT,
Manchester

MS. received 15 *November* 1945

ABSTRACT. Deficiencies of existing theories relating to the propagation of sound along ducts lined with porous sound-absorbing materials are discussed. Attention is drawn to the wave-nature of the propagation of acoustic disturbances in the lining material, and a new theory is put forward in which the propagation of acoustic waves in the composite system of air space and lining is investigated for two simple shapes of duct. The results of the theory are in agreement with those of an earlier theory by Morse as long as the attenuation in the duct is not large, but wide divergences appear, in particular where the attenuation in the duct begins to approach that characteristic of the propagation of sound through bulk lining material. In the latter case it is shown by a particular experimental example that the attenuation found in practice is in agreement with the theory described in this paper.

§ 1. INTRODUCTION

Various aspects of the problem of the transmission of sound through pipes and channels have been investigated in the past by theory and experiment. For pipes with smooth rigid walls, the primary effect of walls is to redirect part of the sound by reflection. Apart from the influence of viscosity on the thin layer of air near the wall, the energy of the waves is conserved during passage along the pipe and the sound is propagated with the velocity characteristic of unrestricted air. Pipes with non-rigid walls or with walls absorbent to sound can affect the free passage of sound waves considerably, and under appropriate circumstances the presence of the walls leads to rapid attenuation of the sound and to substantial modification of its velocity of propagation.

Theories of the absorption of sound waves in pipes lined with absorbing material have been formulated by several authors, with varying success, in terms of the acoustic impedance* of the lining material. For pipes or channels in which, for example, the attenuation is large, this approach is not valid, and in the work which follows, a more complete theory is provided for cases in which the walls are lined with porous material which is homogeneous and isotropic. This theory takes account of propagation of sound in the lining itself.

§ 2. PREVIOUS THEORIES AND THEIR LIMITATIONS

In the past, the behaviour of an absorbing material in a sound field has most frequently been described in terms of the absorption coefficient for normal incidence. This quantity is the fraction of the energy of a plane sound wave incident normally on the surface that is not reflected by the material. The absorption coefficient refers only to the transfer of energy from a wave which falls in a specified direction on a surface and, therefore, conveys no information about the treatment received by waves which are directed obliquely at the surface of the material, as in the case of waves passing through a pipe lined with the material.

---

* The use, as in this paper, of the description " acoustic impedance of the lining " to refer to the unit-area impedance at one surface of the lining when the other surface is backed by a substantially rigid wall, and when sound falls at normal incidence on the surface, is general in the American literature on this subject.

In most recent theoretical studies in acoustics, as, for example, in the theories of Morse (1939 a) and others relating to the decay of sound in rooms, the absorption coefficient has been displaced in favour of the much more informative quantity—the acoustic impedance at the wall. This quantity is the complex ratio of the oscillatory pressure at the surface of the material to the resultant oscillatory velocity at the surface for a plane wave incident normally on the surface. In 1939 Sivian showed that by assuming that the oscillatory pressure across the pipe is uniform, the propagation characteristics of waves travelling down a pipe lined with absorbing material can be expressed in terms of the acoustic impedance of the lining and of the geometrical form and dimensions of the pipe. He compared the attenuation computed from his formula with the attenuation measured in a circular pipe lined with an inch-thick layer of rock-wool. Satisfactory agreement was obtained for low frequencies, but for higher frequencies (over 2000 c/s. for a pipe of radius 10 cm.) the theory failed to explain the comparatively low attenuation found in practice. The cause of this failure is not difficult to find, and lies in the assumption of uniform oscillatory pressure across the pipe. Rayleigh has shown that a general acoustic disturbance travels along a tube as a substantially plane wave only when the radius is less than 0·29 of the wave-length. For a tube with walls of sound-absorbing material, the pressure across the tube is unlikely to be uniform beyond frequencies for which the radius of the tube approaches the quarter wave-length of the sound, since at such frequencies diffraction effects are insufficient to allow uniform diffusion of the sound across the tube.

The main deficiency of Sivian's theory has been removed by Morse (1939 b), who has derived a theory of the attenuation of sound in uniform, rectangular and circular lined ducts for cases wherein the whole effect of the wall-lining on the wave can be described in terms of the acoustic impedance of the lining. Morse has derived expressions for the attenuation constant for each of the possible modes of propagation in the duct and has published comprehensive charts from which to evaluate the attenuation constant. Beranek (1940) has applied Morse's theory to certain cases of square and rectangular ducts lined with rock-wool and has found good agreement with experiment. Rogers (1940) has derived theoretical expressions which are substantially the same as those of Morse and, for the very special case of material possessing only a resistive component of acoustic impedance, has reduced the expressions to simpler terms.

Implicit in Morse's theory is the assumption that the motion at the surface depends only on the acoustic impedance and on the local acoustic pressure and not on the acoustic pressure eleswhere. For homogeneous material of the porous type, this assumption leads to results which frequently are not seriously in error. Nevertheless, the assumption is not strictly tenable, and it is proposed in the present paper to consider in more detail the behaviour of acoustic waves in a lined pipe.

It has been surmised (Crandall, 1927) that sound is propagated as an attenuated wave in a porous material, and recent work (Scott, 1946 a) by the author has demonstrated that this view leads to an accurate description of the acoustic disturbance in the material. Part of the initial energy of a wave which travels along a pipe lined with porous material is, therefore, transmitted through the lining itself. The oscillatory velocity at the surface of the lining depends in consequence on the joint action of the wave disturbances inside and outside the lining, and cannot be expressed merely in terms of the acoustic impedance and the oscillatory pressure at the surface of the lining. In the theory which is given below, the course of the wave is traced through the composite space of air passage and pipe lining combined. The resulting expressions show that, as in Morse's theory, the acoustic disturbance can be resolved into modes which possess fixed propagation constants. Under certain conditions, notably when the attenuation in the duct is small and where the lining consists of a densely packed collection of fibres, the expressions for the propagation constant in the duct tend to the same form as do those of Morse. Where the attenuation is higher, as, for example, in narrow ducts and at high frequencies in the acoustic range, Morse's expression gives attenuations of much higher magnitude than are given by the more complete theory.

§ 3. THE PASSAGE OF ACOUSTIC WAVES ALONG A CHANNEL LINED WITH POROUS MATERIAL

In the theory that follows, the porous lining material is considered as homogeneous and isotropic in so far as concerns the propagation of acoustic waves. For such materials it has been shown, both on acceptable theoretical grounds (Monna, 1938, and Morse and Bolt, 1944) and by experiment (Scott, 1946 a), that any acoustic disturbance in the medium must satisfy a wave equation of simple type.

Thus if the vector **v**, which will be called the velocity of average flow, represents the rate of volume flow through unit area of cross-section normal to **v** in the lining material, it can be shown (see Scott, *loc. cit.*) that the equation for the velocity potential of average flow ψ, defined so that

$$\mathbf{v} = -\operatorname{grad}\psi \qquad \ldots\ldots(1)$$

is

$$\nabla^2\psi + h^2\psi = 0; \qquad \ldots\ldots(2)$$

h is here a complex quantity and is simply related to the attenuation and phase velocity of plane sound waves in the porous medium. Thus

$$jh = \alpha + j\beta,$$

where α is the attenuation constant and β is the wave-length constant of an acoustic wave in the bulk material. The equation for the acoustic pressure in the pores of the porous material can be shown to be given by $p = \rho'\dot{\psi}$, and since the present concern is only with acoustic disturbances of fixed pulsatance ω, this may be written

$$p = j\omega\rho'\psi. \qquad \ldots\ldots(3)$$

ρ' is a second complex quantity which is related to the characteristic impedance of the lining material. The parameters h and ρ' can be obtained experimentally from measurements performed on the material of the lining in bulk (Scott, *loc. cit.*). Equations (1), (2) and (3) together describe the acoustical disturbance in the lining of the channel. Corresponding expressions for the velocity potential of flow ϕ in the air passage are well known to be

$$\mathbf{v} = -\operatorname{grad}\phi, \qquad \ldots\ldots(4)$$

$$\nabla^2\phi + k^2\phi = 0, \qquad \ldots\ldots(5)$$

and

$$p = j\omega\rho_0\phi, \qquad \ldots\ldots(6)$$

where $\mathbf{v}$ is the oscillatory velocity in the air passage, while k is the wave-length constant (2π/wave-length) and ρ_0 is the density of air.

Determination of the nature of waves which travel along the channel involves the solution of equations (1) to (6) consistent with the boundary conditions imposed by the channel.

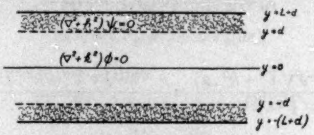

Figure 1. Plan of channel

Figure 1 represents a cross-section of the channel. Suppose the width of the air space is $2d$ and that the linings are each of depth L. Let x represent distances along the channel and y distances away from the central plane of the channel. The equation for ψ in the linings becomes

$$\frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial x^2} + h^2\psi = 0 \qquad \ldots\ldots(7)$$

and

$$p = j\omega\rho'\psi \qquad \ldots\ldots(8)$$

The equations for ϕ in the central air space become

$$\frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial x^2} + k^2\phi = 0 \qquad \ldots\ldots(9)$$

and

$$p = j\omega\rho_0\phi. \qquad \ldots\ldots(10)$$

For any one mode of propagation of a wave along the duct, ϕ and ψ will have a common *propagation constant* Γ such that

$$\phi = \phi(y)e^{-\Gamma x}e^{j\omega t}, \qquad \ldots\ldots(11)$$

$$\psi = \psi(y)e^{-\Gamma x}e^{j\omega t} \qquad \ldots\ldots(12)$$

Substituting these expressions in (7) and (9) and solving, we have:—

$$\phi(y) = Ae^{j\lambda y} + Be^{-j\lambda y} \text{ where } \lambda^2 = \Gamma^2 + k^2,$$

$$\psi(y) = Ce^{j\mu y} + De^{-j\mu y} \text{ where } \mu^2 = \Gamma^2 + h^2,$$

and where A, B, C and D are constants whose values are yet to be found.

When $y=0$, $-\dfrac{\partial\phi}{\partial y}=0$, by symmetry ,

$$\therefore \quad A = B. \qquad \ldots\ldots(13)$$

At the outer wall, $y=L+d$ and the flow velocity normal to the wall $\left(-\frac{\partial\psi}{\partial y}\right)$ is zero;

$$\therefore \quad j\mu Ce^{j\mu(L+d)} - j\mu De^{-j\mu(L+d)}$$

or

$$D = Ce^{2j\mu(L+d)} \qquad \ldots\ldots(14)$$

At the surface of the lining ($y=d$), there must be continuity of the average flow velocity and of the oscillatory pressure, hence:

$$j\lambda Ae^{j\lambda d} - j\lambda Be^{-j\lambda d} = j\mu Ce^{j\mu d} - j\mu De^{-j\mu d}$$

and

$$\rho_0(Ae^{j\lambda d} + Be^{-j\lambda d}) = \rho'(Ce^{j\mu d} + De^{-j\mu d}).$$

Substituting for B and D from (13) and (14) and eliminating A and C,

$$\rho_0 \frac{\coth(-j\lambda d)}{\lambda} = \frac{\rho' \coth j\mu L}{\mu},$$

which may be written

$$\rho_0 \frac{\coth(-j\sqrt{\Gamma^2+k^2}\,.\,d)}{\sqrt{\Gamma^2+k^2}} = \frac{\rho' \coth j\sqrt{\Gamma^2+h^2}\,.\,L}{\sqrt{\Gamma^2+h^2}}. \qquad \ldots\ldots(15)$$

The equation (15), therefore, gives the propagation constant of the mode in terms of the dimensions of the duct and the parameters ρ' and h of the absorbing material.

§ 4. THE PASSAGE OF ACOUSTIC WAVES ALONG A CIRCULAR PIPE LINED WITH POROUS MATERIAL

Expressions can be obtained in a manner similar to that used above for the propagation constant associated with a mode of propagation along a circular pipe lined with a layer of uniform thickness of homogeneous, isotropic porous material. Figure 2 shows the cross-section of a pipe of radius b lined with a thickness $(b-a)$ of porous material. Equations (1), (2) and (3) again give the velocity potential of average flow in the material of the lining, and (4), (5) and (6) give the velocity potential of flow in the air passage. Let z represent distances along the axis of the pipe and let r represent radial distances from the axis. Radial symmetry allows us to write

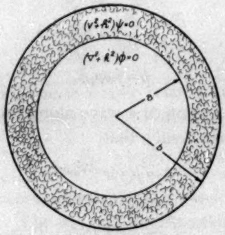

Figure 2. Section of tube

$$\nabla^2 = \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2}\right).$$

If Γ again represents the propagation constant of any one mode, the velocity potentials may be written:

$$\phi=\phi(r)e^{-\Gamma z}e^{j\omega t}, \qquad \ldots\ldots(16)$$

$$\psi=\psi(r)e^{-\Gamma z}e^{j\omega t}, \qquad \ldots\ldots(17)$$

whence, by substitution in (2) and (5),

$$\frac{\partial^2\phi}{\partial r^2}+\frac{1}{r}\frac{\partial\phi}{\partial r}+\lambda^2\phi=0, \text{ where } \lambda^2=\Gamma^2+k^2,$$

$$\frac{\partial^2\psi}{\partial r^2}+\frac{1}{r}\frac{\partial\psi}{\partial r}+\mu^2\psi=0, \text{ where } \mu^2=\Gamma^2+h^2.$$

The solutions for $\phi(r)$ and $\psi(r)$ are, therefore,

$$\phi(r)=AJ_0(\lambda r)+BY_0(\lambda r), \qquad \ldots\ldots(18)$$

$$\psi(r)=CJ_0(\mu r)+DY_0(\mu r). \qquad \ldots\ldots(19)$$

The boundary conditions are as follows:

$$r=0, \quad \phi \text{ finite, whence } B=0,$$

$$r=a, \quad -\frac{\partial\phi}{\partial r}=-\frac{\partial\psi}{\partial r}, \quad \text{and}$$

$$j\omega\rho_0\phi=j\omega\rho'\psi$$

$$r=b, \quad -\frac{\partial\psi}{\partial r}=0.$$

Hence, by differentiation, etc., of (18) and (19),

$$\lambda AJ_1(\lambda a)=\mu\{CJ_1(\mu a)+DY_1(\mu a)\} \qquad \ldots\ldots(20)$$

$$\rho_0 AJ_0(\lambda a)=\rho'\{CJ_0(\mu a)+DY_0(\mu a)\} \qquad \ldots\ldots(21)$$

and $\quad 0=\mu\{CJ_1(\mu b)+DY_1(\mu b)\},$

From the latter equation $D=-CJ_1(\mu b)/Y_1(\mu b)$, whence, by substitution in (20) and (21) and then eliminating A and C we find,

$$\frac{\rho_0}{\lambda}\frac{J_0(\lambda a)}{J_1(\lambda a)}=\frac{\rho'}{\mu}\,\frac{J_0(\mu a)Y_1(\mu b)-Y_0(\mu a)J_1(\mu b)}{J_1(\mu a)Y_1(\mu b)-Y_1(\mu a)J_1(\mu b)},$$

and since $\lambda^2=\Gamma^2+k^2$ and $\mu^2=\Gamma^2+h^2$, the equation for the propagation constant Γ becomes

$$\frac{\rho_0}{\sqrt{\Gamma^2+k^2}}\,\frac{J_0(\sqrt{\Gamma^2+k^2}\,.\,a)}{J_1(\sqrt{\Gamma^2+k^2}\,.\,a)}=\frac{\rho'}{\sqrt{\Gamma^2+h^2}}$$

$$\times\,\frac{J_0(\sqrt{\Gamma^2+h^2}\,.\,a)Y_1(\sqrt{\Gamma^2+h^2}\,.\,b)-Y_0(\sqrt{\Gamma^2+h^2}\,.\,a)J_1(\sqrt{\Gamma^2+h^2}\,.\,b)}{J_1(\sqrt{\Gamma^2+h^2}\,.\,a)Y_1(\sqrt{\Gamma^2+h^2}\,.\,b)-Y_1(\sqrt{\Gamma^2+h^2}\,.\,a)J_1(\sqrt{\Gamma^2+h^2}\,.\,b)}. \qquad \ldots\ldots(22)$$

§ 5. DISCUSSION OF THEORETICAL RESULTS

Equations (15) and (22) give series of values of Γ which satisfy the boundary conditions for, respectively, a channel and a pipe lined with porous material. Each value of Γ is associated with a corresponding variation of ϕ and ψ across the channel as given by (11) and (12) or by (16) and (17), and, therefore, refers to a particular "mode of propagation" of sound in the sytsem. An arbitrary acoustic disturbance at the beginning of the channel or pipe can be resolved into a number of these component modes and the separate modes regarded as travelling along the tube, each with a wave-front of fixed shape, and with fixed attenuation constant and phase velocity.

Morse's theory relating to the propagation of sound in ducts leads to a similar multiplicity in values of Γ and, therefore, to the existence of many modes of propagation for a given duct. Morse's theory, which would have held for the case of the channel lined with porous material if the motion at the surface of the lining had been always normal to the lining, and dependent solely on the oscillatory pressure near the lining, gives as the equation for Γ

$$\rho_0 \frac{\coth(-j\sqrt{\Gamma^2+k^2}\,.\,d)}{\sqrt{\Gamma^2+k^2}} = \frac{Z}{\omega}, \qquad \ldots\ldots(23)$$

where Z is the acoustic impedance of the lining. It can readily be shown (Scott, 1946 b) that Z may be expressed in terms of ρ', h and L in the following manner :

$$Z = \frac{\rho'\omega}{h}\coth jhL. \qquad \ldots\ldots(24)$$

Morse's equation for Γ may therefore be rewritten in a form which compares easily with that obtained in this paper; thus (23) becomes

$$\rho_0 \frac{\coth(-j\sqrt{\Gamma^2+k^2}\,.\,d)}{\sqrt{\Gamma^2+k^2}} = \rho' \frac{\coth jhL}{h}. \qquad \ldots\ldots(25)$$

The two expressions (15) and (25) lead to the same values of the propagation constant Γ for cases in which Γ^2 is very small compared with h^2, that is, when the attenuation constant and the wave-length constant associated with the mode are much less than the attenuation constant and wave-length constant characteristic of the material of the lining in bulk. Low values of attenuation constant are to be found in practice for wide air passages ($2d > 15$ cm. for linings of rock-wool). For smaller values of d, Morse's formula gives attenuations which increase without limit as d is reduced. Such high attenuations are not found in practice, nor are they to be expected, for it is clear that if the air-passage between two linings of porous material is reduced sufficiently, the passage of an acoustic disturbance along the channel is tantamount to the propagation of waves through a channel filled with the lining material. In this case the value to be expected for Γ is that of the propagation constant jh of the porous material. An experimental illustration of this point is given by the data of figure 3. This figure shows measured values of attenuation (in decibels/cm.) determined for a channel with an air-space 1·27 cm. wide flanked by two 5·08-cm. thick linings of rock-wool (stillite of bulk density 0·08 gm./cm.^{3}). Curve A is the measured attenuation, B is the attenuation in the bulk material and C is the attenuation associated with the lowest mode of propagation calculated from Morse's formula. The latter calculation is facilitated by the charts which are published in Morse's paper. The figure also shows the corresponding velocities of propagation of waves in the channel. Thus E is the measured velocity, F is the velocity of a wave in the porous material in bulk and G is the velocity calculated from Morse's formula for the fundamental mode. The experimental results were obtained under conditions in which the disturbance in the channel consisted substantially of the fundamental mode of propagation in the channel. The material used for the linings was closely similar to that used by the author in a recent paper upon the propagation of sound in porous media,* and the experimental values of the parameters h and ρ' therein contained are, therefore, relevant to this material.

The complete solution of equation (15) or of equation (22) involves a process of successive approximation, and in general is extremely tedious; only in exceptional cases is the considerable work of computation justified in practice. It is therefore important to investigate the conditions under which Morse's formula gives a value of Γ that for practical purposes is a sufficiently good approximation to the true value. The difference between the two expressions depends upon the difference between $(\Gamma^2+h^2)^{\frac{1}{2}}$ and h, and for this difference to be small the real and imaginary components of Γ^2 must respectively be much less than the real and imaginary components of h^2. Thus if α_1 and β_1 represent the attenuation constant and wave-length constant corresponding to Γ, i.e. $\Gamma=\alpha_1+j\beta_1$, and if, as before, $jh=\alpha+j\beta$,

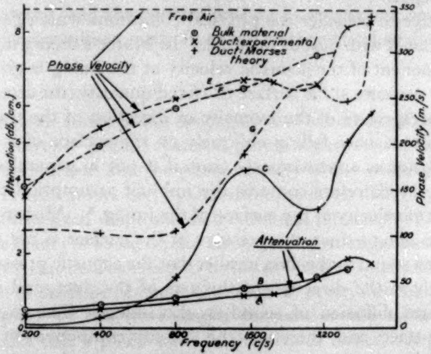

Figure 3. Comparison of results of experiment and Morse's theory for a narrow-lined duct.

$$\Gamma^2=(\alpha_1{}^2-\beta_1{}^2)+2j\alpha_1\beta_1$$

and
$$h^2=(\beta^2-\alpha^2)-2j\alpha\beta,$$

and, therefore, $(\alpha_1{}^2-\beta_1{}^2)$ and $\alpha_1\beta_1$ must be respectively much less than $\beta^2-\alpha^2$ and $\alpha\beta$. For most practical sound absorbents of the porous type, β and α are of the same order of magnitude at all frequencies, and in this case it is necessary that neither α_1 nor β_1 be comparable in magnitude to α or β. Thus, for Morse's expression to apply, the attenuation calculated therefrom must be substantially less than that characteristic of the bulk material, and the velocity of propagation of the sound (of a given mode) must be much higher than that characteristic of the bulk material. Satisfaction of the first condition implies, in general, either that the quantity $Z/\rho c$ be large compared with unity, i.e. the impedance Z must be high compared with the characteristic impedance of air ($\rho_0 c$), or that the duct width be large compared with the wave-length. The existence of low attenuation in the duct does not, however, insure that the second condition be satisfied. Where the ratio of the wave-length constant β of the bulk material to the wave-length constant k of unrestricted air is not very different from unity, β_1 is necessarily of the same order as β, and in consequence Morse's formula again fails. This state of affairs is instanced by the results in the lower frequency range of figure 3. In the case considered, the velocity of propagation in the bulk material is only about half that in free air, i.e. $\beta \simeq 2k$. The value of velocity calculated from Morse's formula is lower than that in the bulk material, and in consequence this calculated value of β_1 is substantially greater than β, which is inadmissible if the results of Morse's formula are to be in agreement with those deduced from the more exact formula of the present paper.

---

* For the data relating to the material used, see tables 2 and 3 for h and ρ' and curves marked 2 in figure 6 for the acoustic impedance Z of the lining.

It is instructive to consider the physical conditions which govern the sound field in the duct lined with porous material. In Morse's theory it is assumed that the normal component of the acoustic velocity at the lining is given by the ratio of the oscillatory pressure at the surface to a fixed quantity (the acoustic impedance of the lining), irrespective of the intensity or direction of the wave in the duct. Whilst for a uniform wave falling obliquely on the surface of a densely packed material this relation is approximately true, it is not in general strictly tenable, and Morse's theory therefore contains the implicit assumption that there is no axial component of velocity at the surface of the lining.* Moreover, for a sound wave propagated along a duct, the pressure at the surface is not uniform. The attenuation of the sound in the duct implies that the acoustic pressure in the lining falls continuously in the direction of the axis of the duct, and in consequence there is a forward diffusion of sound in the lining. The pressure gradient normal to the surface, and, therefore, the velocity component in this direction, is thereby dependent on the pressure at adjacent sections along the duct, and not solely on the local pressure as is required in Morse's theory.

The discrepancy involved in the application of Morse's theory to a duct lined with porous material can be related in a general way to the attenuation (α) and wave-length constant (β) associated with the propagation of uniform waves in the material. The extent to which the pressure gradient normal to the surface of the lining is affected by the acoustic disturbance at preceding sections depends firstly on the direction of the oscillatory velocity in the lining. In the air-space, the velocity vector is mostly longitudinal, but refraction of the wave at the lining tends to turn the vector towards the wall. The "refractive index" at the boundary is clearly given by the ratio β/k of the wave-length constants in the two media.

---

* The author has shown by means of auxiliary experiments that the velocity in the air-space is closely axial even at the boundary surface. For this purpose a shallow duct was built with lined walls contained between glass sheets, one-half inch apart, which served as top and bottom of the duct. Sound was led into the duct at one end and the field, delineated by illuminated smoke-particles, was explored with the aid of a low-power microscope. Over a wide range of width of air-space and duct-lining, and for all frequencies between 100 and 1000 c./s. it was estimated that the inclination of the particle traces to the axis did not exceed 5° at the boundary. See Morse and Bolt (1946), p. 92.

Where this ratio, and, therefore, the turning action are large, the pressure at a point immediately behind the lining surface is not as much affected by the wave at preceding sections as where the ratio is small. The gradient of pressure normal to the lining is, therefore, less dependent on the disturbance at adjacent sections in the former case than in the latter, and the application of Morse's theory is likely to be least in error where β/k is greatest.*

The disturbing effect on the normal gradient of pressure, due to waves falling on preceding sections of the duct, is clearly also dependent on the attenuation coefficient α of the lining material, since where α is high the magnitude of the disturbance at one section due to waves at an earlier section is correspondingly reduced. The application of Morse's theory is less likely to be in serious error where the density of packing of the material is great, and α is high, than where the material is loosely packed.

The above physical considerations illustrate in a qualitative manner the conditions under which Morse's analysis might be expected to be valid for ducts lined with porous material, and the conclusions are in satisfactory agreement with the deductions which have been made from the mathematical analysis of the present paper.

§ 6. CONCLUSIONS

The theory given above of the propagation of sound between walls lined with homogeneous, isotropic porous material provides expressions from which can be calculated the attenuation and the phase velocity of any of the possible modes of propagation. These modes are, by definition, acoustic disturbances in which the wave-front retains a fixed shape and distribution of intensity in its passage through the duct. The theory is based on the assumption that any acoustic disturbance in the porous lining can be described in terms of a wave equation analogous to that which is known to describe acoustic disturbances in air, but containing a complex parameter in place of the wave-length constant of the latter. The complex parameter implies that a plane wave in the porous material is propagated as an attenuated wave, and the real and imaginary components of this parameter are readily deducible from the experimentally determinable attenuation constant and wave-length constant characteristic at any given frequency of plane waves in the bulk material.

The problem of deducing expressions for the attenuation and wave-length constants of any given mode in the duct reduces to the simultaneous solution of the wave equations for the air-space and for the lining, subject to the appropriate boundary conditions dictated by the geometry of the duct. Solutions to this problem are provided for the cases of a channel and a circular pipe lined symmetrically with porous material. Numerically these expressions can be solved in the general case only with difficulty. A number of important deductions are, however, readily made. Thus it is shown in the paper that the expression for the propagation constant is, under special conditions, substantially in agreement with that deduced earlier by Morse on the assumption that the effect of the linings on the passage of sound through the duct can be fully described in terms of the acoustic impedance of the lining. For ducts lined with densely packed sound-absorbing material, and for which the duct width is not small compared with the wave-length of sound in free air, Morse's formulae give results which are in reasonable agreement with those deducible from the formulae of the present paper. Numerical solutions of Morse's formulae can be obtained without undue difficulty from the comprehensive set of charts provided in Morse's paper on this subject (*loc. cit.*).

---

* For commercially available sound-absorbent materials, β/k is rarely very large, except for very dense materials and at frequencies low in the acoustic range. Thus experiments show that for very densely packed rock-wool, the phase velocity in the material is about 40 m/s. at 100 c./s., which implies a refractive index of 8·6 at the boundary surface. The refractive index for commoner materials usually ranges from 1 at high frequencies to 3 at low frequencies.

Where the sound-absorbing lining consists of a loosely packed material, or where the width of the air-space in the duct is small compared with the wave-length, Morse's formulae no longer give a reasonably good approximation to the proper solution of the problem. Thus, in particular, the attenuation constant calculated from Morse's formulae is infinite for the case of a vanishingly small air-space, whereas it is clear from general physical considerations, as also from the expressions deduced in the present paper, that in this case the acoustic disturbance is propagated with the attenuation constant characteristic of the propagation of sound in the porous material of the lining. Experimental results are included in the paper to illustrate this point.

§7. ACKNOWLEDGMENT

The author wishes to express his thanks to Dr. A. J. King for his interest in this paper, to Mr. B. G. Churcher, Manager of the Research Department, for his authorization of the work, and to Sir Arthur P. M. Fleming, C.B.E., Director of Metropolitan-Vickers Electrical Co., Ltd., for permission to publish this paper.

REFERENCES

Beranek, L. L., 1940. *J. Acoust. Soc. Amer.*, **11**, 228.
Crandall, I. B., 1927. *Vibrating Systems and Sound*, p. 197 (London: Macmillan).
Monna, A., 1938. *Physica*, **5**, 129.
Morse, P. M., 1939 a. *J. Acoust. Soc. Amer.*, **11**, 56.
Morse, P. M., 1939 b. *J. Acoust. Soc. Amer.*, **11**, 205.
Morse, P. M. and Bolt, R. H., 1944. *Rev. Mod. Phys.*, **16**, 69.
Rogers, R., 1940. *J. Acoust. Soc. Amer.*, **11**, 480.
Scott, R. A., 1946 a. *Proc. Phys. Soc.*, **58**, 165; 1946 b. *Ibid.*, **58**, 253.
Sivian, L. J., 1939. *J. Acoust. Soc. Amer.*, **11**, 56.

CATHODO-LUMINESCENCE: PART I. GROWTH AND DECAY PROCESSES

By J. W. STRANGE and S. T. HENDERSON,
Electric and Musical Industries, Ltd., Hayes, Middlesex

MS. received 10 *September* 1945

ABSTRACT. The active centre theory of luminescence, with certain assumptions, leads to the expectation of exponential growth and decay processes in light emission. Measurement has shown the widespread occurrence of exponential processes, though the details are much more complex than the simple theory would suggest. The effects of changing current density and temperature were examined. The phosphors studied fell into two classes, distinguished mainly by the rates of the processes measured. Some non-exponential processes were found, especially in long time decay, but none were definitely bimolecular.

§ 1. INTRODUCTION

The work described in this and the following paper was mostly done in 1937–1940. References to parts of it have been published elsewhere (Strange, 1938, 1940). It is concerned with the phenomena of light emission in luminescent materials under cathode-ray bombardment, especially with the very rapid changes which occur during this process and with the changing efficiencies observed as conditions are varied. Previous work has largely concentrated on excitation by ultra-violet radiation, and slow phosphorescent decays have been studied far more than the more rapid processes.

It is desirable to define some terms that will be used, though the outline of the *active centre* theory of luminescence will be assumed already known. A *phosphor* or luminescent material means here a solid inorganic base material which has been heat-treated, generally with a small addition of some foreign metal or its salt (the *activator*) in such a way that the product emits visible light under a beam of cathode rays, ultra-violet or other radiation. The mechanism is a process of raising the atoms of the phosphor to higher energy states. Their return with the emission of light (*luminescence*) may be classified as *fluorescence* or *phosphorescence* according to the characteristics observed, the former being comparatively rapid and due to electrons making inner transitions, atomic or molecular, and the latter being temperature dependent owing to the metastable states involved (Pringsheim, 1939). The word *excitation* must be restricted to the process by which the impinging radiation raises electrons from ground states to higher levels of energy. The light emission at any moment depends on the number of electrons in excited states and on the probability of their return to ground states: the emission increases from zero when excitation begins, but attains its final value only after a finite time. This is the *growth process.* When light output has thus become constant under steady excitation, the *equilibrium state* has been reached. After the exciting radiation has been cut off, the light emission declines to a low value in a time generally comparable with that taken in the growth process; this is called the *decay process.*

We may refer briefly to Becquerel's work (1861). He demonstrated the rapid exponential decay with time of the light emitted by uranyl salts (fluorescence), while the slow phosphorescent decay of zinc sulphides gave the relation

$$cI_0^m = I^m(t+c), \qquad \ldots\ldots(1)$$

where c and m are constants, I_0 and I the intensity of light emitted respectively at the start and after a decay time t. If $m=\frac{1}{2}$, the usual bimolecular form results:

$$I_0 = I(1+\gamma t)^2. \qquad \ldots\ldots(2)$$

Lenard (1928) attempted to explain all long-period decays as due to the sum of several exponential processes with different constants. This is an unsatisfactory way of interpreting an experimental curve, and to a limited extent the same difficulty occurred in the present work.

Among the few modern papers giving results of cathodo-excitation of phosphors are those of Schleede and Bartels (1938), and Nelson, Johnson and Nottingham (1939). De Groot's work (1939), though concerned with ultra-violet excitation, is of considerable interest.

From the active-centre theory may be derived expressions for light output during and after the excitation period, if certain additional assumptions are made. The treatment given is due to Dr. L. F. Broadway, and it was the original motive for the experimental work described below. A similar theoretical deduction has since been published by Nelson, Johnson and Nottingham (1939). No use has been made of recent work on trapping states, published since this paper was written (Randall, Wilkins, Garlick, 1945), but this new work does not appear to afford so direct an approach to the problem as the discussion which follows. See also the introduction to Part III of the present series.

§ 2. THEORETICAL

(a) *Monomolecular or random type of process*

Consider the excitation of a thin slab of the phosphor, of unit area and thickness δy, at a depth y from the surface, and having n active centres per unit volume, of which number x are in excited states at time t from the start of excitation. The slab is assumed to be under a constant beam of i electrons per unit area per unit time, and the incident voltage of these electrons, V, is assumed constant (i.e., the straggling and scattering effects when $y>0$ are neglected). Two probability constants are introduced: A for the excitation of centres by incident electrons at voltage V, and B for the return of excited electrons to lower states with emission of radiation. A must vary with y, but is here assumed constant, since δy is very small. A and B are both taken to be independent of i.

Then in time δt the number of active centres excited in unit volume is

$$\delta_1 x = Ai(n-x)\delta t.$$

The number of centres decaying with radiation in the same time is $\delta_2 x = Bx\,.\,\delta t$. Hence the net number of centres excited is

$$\delta_1 x - \delta_2 x = \{Ai(n-x) - Bx\}\delta t,$$

and the "growth process" is given by

$$\frac{dx}{dt} = Ain - x(Ai - B)$$

or $$x = Ain\{1 - \exp[-(Ai + B)t]\}/(Ai + B). \qquad \ldots\ldots(3)$$

Thus at time t the light output in quanta/unit vol./unit time from the thin slab of phosphor is

$$L = Bx = AB\,in\{1 - \exp[-(Ai + B)t]\}/(Ai + B), \qquad \ldots\ldots(4)$$

which requires an exponential rise of L as the excitation continues, with a *growth constant* $Ai + B$. At the end of excitation the constant B is alone concerned, and an exponential decay occurs, with *decay constant* B.

The total light output from a slab of finite thickness, or $\int L\,.dy$, neglecting the absorption and scattering of light and the possible decrease of i as electrons penetrate the slab, will be assumed to be of the same form as equation (3) for the experimental conditions used, namely, a constant incident voltage at $y = 0$ and, complete absorption of the beam, which implies that the value of A applicable to the whole slab of phosphor bears a simple relation to A used above. The new value of A is denoted by A'.

At equilibrium excitation, where t is large, the light output becomes

$$L = A'B\,in/(A'i + B), \qquad \ldots\ldots(5)$$

from which it appears that L is proportional to the current density when this is small, while at very high values of i the light output reaches a constant value Bn.

(b) *Bimolecular type*

If the active centre is ionized, the rate of recombination or decay, and hence the light output ,will be proportional to Bx^2 instead of to Bx, and by an argument similar to the above, the light output L at equilibrium excitation is given by

$$L(2B) = 2A'B\,in + (A'i)^2 - A'i\sqrt{\{(A'i)^2 + 4A'B\,in\}}. \qquad \ldots\ldots(6)$$

Again L is proportional to the current density when this is small. If there are x_0 excited centres at the end of excitation, it is readily shown that, at low current densities,

$$x_0 = \sqrt{(A'in/B)}.$$

With the usual symbols,

$$-\frac{dx}{dt} = Bx^2 = I.$$

Hence $I_0 = I(Bx_0t + 1)^2$, and from equation (2) $\gamma = \sqrt{(A'B)}$ at low current densities, when it also follows that the growth curve is at first convex towards the time axis.

§3. EXPERIMENTAL METHODS AND RESULTS

(a) *Processes lasting up to* 1 *millisecond*

The apparatus consisted of a master oscillator generating frequencies from 500 to 2500 c.p.s., which controlled a pulse generator giving a train of rectangular pulses of about 100 v. amplitude over the same frequency range. These pulses applied to the modulator of the electron gun in the cathode-ray tube switched the electron beam on and off at the frequency required, giving pulses of light from the luminescent screen. A final anode voltage of 5 kv. was used in all the work to be described, except where otherwise stated.

The on and off periods were originally equal, but later the off period was increased to lessen the effect of incomplete decay of the phosphor sometimes observed.

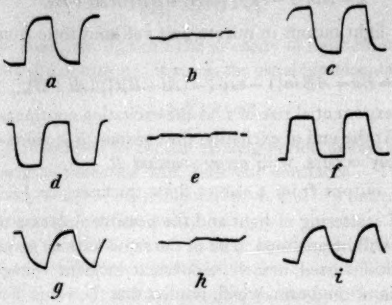

Figure 1. Null method of estimating growth and decay constants. Abscissae represent time; growth or decay period = 500 μ sec.

a. Photocell output, ZnS . Ag, 12 μa./cm.^{2}
b. ,, ,, cancelled.
c. Cancelling signal.
Similarly *d, e, f* for ZnS . Ag, 180 μa./cm.^{2}
g, h, i for CaO . Sm, 10 μa./cm.^{2}

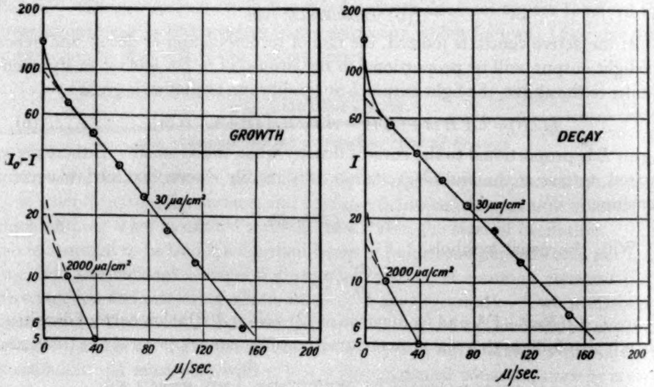

Figure 2. Growth and decay for ZnS . Ag: semi-logarithmic plots. Light units are arbitrary and different for different current densities.

The light emitted was received by a photocell and amplifier, and the amplified signals applied to one pair of plates in a 3-inch electrostatic cathode-ray monitor tube. Both parts of this recording end of the apparatus had an overall frequency response which was flat up to one megacycle per second, a necessary condition to avoid "aperture" effects. The cathode-ray tube with the test samples was of demountable construction to provide for rapid changes of phosphors during operation and of the electron gun when required.

The apparatus described represented the growth and decay curves on a linear time base as shown in figure 1, *a*, *d*, *g*. These were photographed at different current densities and worked out graphically to give growth and decay curves such as figure 2 for ZnS . Ag. The fundamental exponential character is clear.

Later improvements extended the range of working down to currents of 1 μa./cm^2, and a null method was also developed which allowed the feeding-in of one or more exponential voltage signals, derived by charge or discharge of condensers, to the recording monitor, so as to cancel the original trace; growth and decay could be separately treated if required, and the exponential time constants thus determined much more readily than by the photographic methods, and at least as accurately. Perfect cancellation was not usually obtained, but figure 1 shows some typical results. The left-hand column gives the cathode-ray tube trace produced by the photocell output for the material and current density indicated, the right-hand column the corresponding cancelling signals as recorded without the photocell output, and the centre column gives combinations of the two.

General survey of Ag class phosphors

It appeared at once that the lowest available frequency was too fast to give any results from manganese activated materials (see § 3 (b) below), while many oxides, sulphides, tungstates and silicates with other activators (or none) gave similar results to ZnS . Ag with the present apparatus, and the term *Ag class* was adopted to include all such phosphors. At the lowest current densities, the brightest of these gave non-exponential processes, possibly bimolecular. In the more easily accessible range, exponential growth and decay of similar time constants were found quite generally, with a faster initial exponential process in each (figure 1), which increased in relative proportion with increase of current density. The reservation already made regarding the validity of sums of exponentials must be borne in mind in examining table 1, which gives a general survey of decay constants found at a current density of 5 μa./cm^2 (except for CaO . Sm at 10 μa./cm^2). The fast and slow processes are named α and β respectively. Complex emission spectra (bands and fairly sharp lines) of the samarium-activated phosphors were here treated as a whole, but for materials described in section 3 (b) it was sometimes possible to separate out the decay processes for different parts of the emission spectrum.

ZnS . Ag, the material most used in this work, and typical of the class, contained 0·01% Ag and was activated with a chloride flux at about 900° c.

The noteworthy features of these results are: the very rapid decay of zinc oxide; the similarity of the tungstates and, in a different range, of the zinc sulphides; and the independence of these measured processes with respect to long-time phosphorescence, such as that induced by lead in tungstates, or by copper in zinc sulphides. Further, it was shown that ZnS . Ag containing up to one part in a million of nickel suffered no appreciable changes in these decay constants, although the phosphorescence was drastically reduced by the nickel. This supports the view that while both ionization and random (i.e. exponential) processes may occur in these materials, they are acting in parallel and not consecutively.

Table 1. Decay constants and proportions of processes

Material	Constant of α process (sec.$^{-1}$)	Constant of β process (sec.$^{-1}$)	% α	% β
ZnO	$1\,.\,10^{6}$	—	95	5
ZnSe	$1\,.\,10^{6}$	—	90	10
$CaWO_4$. Sm	$1\,.\,10^{6}$	$7\,.\,10^{4}$	5	95
$CaWO_4$	$1\,.\,10^{6}$	$6\,.\,10^{4}$	40	60
$CdWO_4$	$1\,.\,10^{6}$	$6\,.\,10^{4}$	40	60
$ZnWO_4$	$1\,.\,10^{6}$	$6\,.\,10^{4}$	20	80
$ZnWO_4$. Pb	$1\,.\,10^{8}$	$6\,.\,10^{4}$	20	80
SrS . Bi	$2\,.\,10^{5}$	$3{\cdot}5\,.\,10^{4}$	60	40
SrS . Ag	$1\,.\,10^{5}$	$1{\cdot}4\,.\,10^{4}$	60	40
CaS . Cu	$1\,.\,10^{6}$	$2\,.\,10^{4}$	70	30
ZnS . CdS . Cu (45 % Zn . S)	$1\,.\,10^{5}$	$1{\cdot}5\,.\,10^{4}$	30	70
ZnS	$1\,.\,10^{5}$	$1{\cdot}3\,.\,10^{4}$	50	50
ZnS . Cu	$1\,.\,10^{5}$	$1{\cdot}3\,.\,10^{4}$	50	50
ZnS . Ag	$1\,.\,10^{5}$	$1{\cdot}3\,.\,10^{4}$	50	50
ZnS . Pb	$1\,.\,10^{5}$	$1{\cdot}3\,.\,10^{4}$	50	50
ZnS . Au	$1\,.\,10^{5}$	$1{\cdot}3\,.\,10^{4}$	50	50
Zn_2SiO_4	$7\,.\,10^{4}$	$1{\cdot}0\,.\,10^{4}$	40	60
Zn_2SiO_4 . Ag	$7\,.\,10^{4}$	$1{\cdot}0\,.\,10^{4}$	40	60
CaO . Sm	—	$9{\cdot}5\,.\,10^{3}$	5	95

Temperature effects

Other evidence on the nature of the processes was provided by measurements at low and high temperatures. For low temperatures, small sealed-off cathode-ray tubes were tested at 20° c. and again when the screen end was immersed in liquid oxygen. For both ZnS . Ag and $CdWO_4$ at 100 μa./cm.^{2} the α and β decay constants were unchanged by cooling, but a small increase in the proportion of the faster process occurred for each at $-180°$ c. Observations above room temperature were made on a demountable tube with the screen end heated in a tubular oven. Up to 150° c. ZnS . Ag showed a gradual increase of growth and decay " constants " up to twice their original values, and ZnS . Cu gave a steady decay constant up to 150° c., with approximate doubling of its value between 150° c. and 250° c. It is doubtful if the variations are really significant in view of the widely changing efficiencies of the phosphors with temperature, and the much greater variations with current density described below. We may say that the rates of growth and decay are not profoundly altered by temperature changes, in agreement with the view that random electron transitions are responsible for these processes.

Current density effects

Apart from changes in proportions of fast and slow decay processes, indications were soon met that current density changes also affected the values of the " constants" of growth and decay. Table 2 gives decay figures for ZnS.Ag. The growth figures were similar in trend, with somewhat greater values for the constants.

Table 2. Decay of ZnS.Ag

Constants of processes			Percentages of processes	
μa./cm.^{2}	α (sec.$^{-1}$)	β (sec.$^{-1}$)	α	β
5	$1 \, . \, 10^5$	$1{\cdot}3 \, . \, 10^4$	50	50
12	$1 \, . \, 10^5$	$1{\cdot}45 \, . \, 10^4$	63	37
60	$1 \, . \, 10^5$	$1{\cdot}65 \, . \, 10^4$	75	25
180	$1{\cdot}2 \, . \, 10^5$	$4{\cdot}1 \, . \, 10^4$	87	13
700	$1{\cdot}8 \, . \, 10^5$	$5{\cdot}0 \, . \, 10^4$	90	10
5000	$4{\cdot}6 \, . \, 10^5$	$1{\cdot}0 \, . \, 10^5$	95	5

From the above results we are justified in assuming the widespread occurrence of exponential processes in this class of phosphors, but it is clear that the simple theory is far from accounting for the details. The other class of phosphors, the *Mn class*, will now be considered.

(b) *Processes lasting from* 1 *millisecond to* 1 *second*

Another circuit, based on a similar idea to that employed in the measurement of the constants of the Ag class, was used for this part of the work. The general layout of the apparatus is shown in figure 3, and a description is in the Appendix (p. 382).

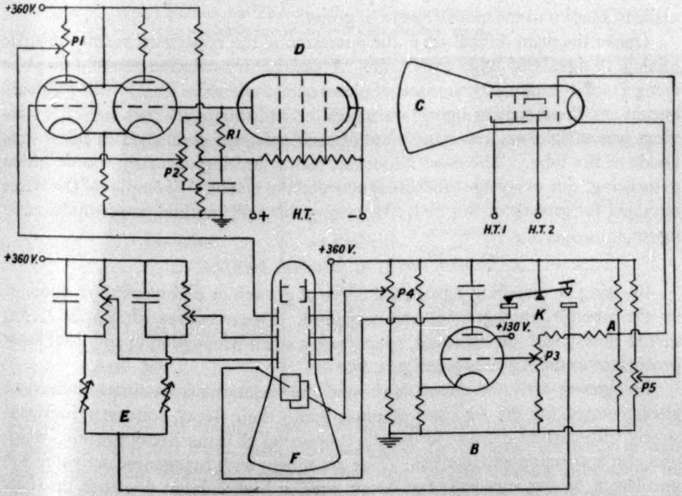

Figure 3. Luminescence decay apparatus.

The depression of the key K started simultaneously the following four processes:—(*a*) the scanning of the monitor tube, (*b*) the decay of the screen under test after equilibrium excitation, (*c*) the rise of potential in the output stage of the photoelectric unit due to decrease in the luminescent light, and (*d*) the exponential fall in potential across the condensers of the matching unit, balanced against the rise in (*c*).

The operation of the apparatus consisted of adjusting the current in the tube to the required value, interposing filters between the screen and the photocell unit to bring the output signal within the linear range of the amplifier, and then matching this signal with one or more signals from the matching unit, varying them in amplitude and constants until an accurate null value was obtained.

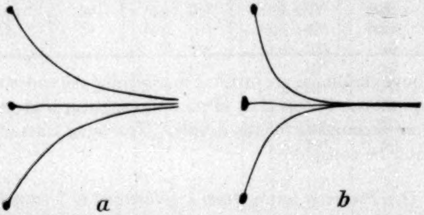

Figure 4. Growth process for Zn_2SiO_4.Mn. Abscissae represent time; total period shown $\simeq$ 40 m.sec. (*a*) 32 μa./cm.² ; (*b*) 200 μa./cm.² Photocell output below, signal from matching unit above, mixture central in each case.

The circuit given is for the measurement of decay, but simple alterations can be made to adapt it to the measurement of growth.

One of the main difficulties in the operation of the apparatus was the accurate mixing of signals of large amplitude. A satisfactory solution was obtained by using a monitor tube with four sets of plates; the three pairs used for the photocell output and the matching signals were parallel, and the fourth pair, used for scanning, was at right angles. One plate of each pair was connected to the second anode of the tube. The most important factor limiting accuracy was deflexion defocusing, due to voltage differences between the plates. Examples of the traces obtained for growth of Zn_2SiO_4.Mn, using this null method, are reproduced in figure 4.

General survey of Mn class phosphors

By using no matching signals, the curve of growth or decay could be obtained on the monitor, photographed and measured. Decay curves for Zn_2SiO_4.Mn agreed closely with exponentials, apart from a small proportion of an initial faster process, as in the Ag class (see figure 5).

Manganese-activated materials showed the variations from simple behaviour already noted for the Ag class of phosphors; their decay constants increased slowly with current density, as did the proportion of faster decay process, which was also apparently exponential. Zinc aluminate with manganese activator was peculiar in having very little fast decay, even at high current densities, and later work on Al_2O_3.Cr (see opposite page) afforded a similar case.

The willemite used was Zn_2SiO_4 with 0·3% Mn. The ZnS . Mn contained 0·1% Mn and was activated at about 1100° c.

The nature of the growth process differed from the simple theory even more than did the decay process. The growth constant was about the same as the decay constant at very low current densities, but increased rapidly with current density, though scarcely at all in the linear manner to be expected (equation (4)). Figure 6 shows the changes for Zn_2SiO_4 .Mn. The growth processes consisted of rapid and slow parts, though the constant given refers to the slow β part. The values for

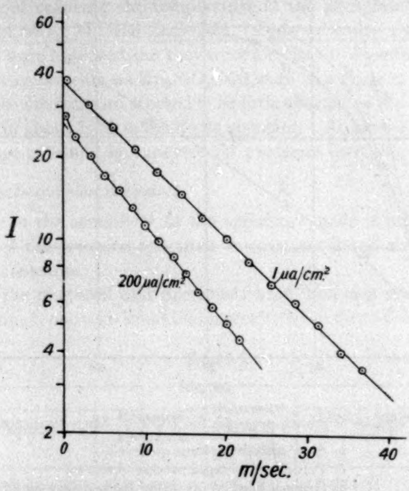

Figure 5. Decay process for Zn_2SiO_4.Mn : semi-logarithmic plots. Light units are arbitrary, and different for the two cases.

Table 3. Decays in manganese class

Phosphor	μa./cm.²	β constant (sec.$^{-1}$)	% α process
$Zn_2B_2O_5$. Mn	1–300	63– 76	2–12
Zn_2SiO_4 . Mn	1–200	67– 78	0–23
ZnO . BeO . SiO_2 . Mn	1–400	98–123	8–25
$ZnAl_4O_7$. Mn	2–900	130–144	0– 4
ZnS . Mn	2–600	625–645	0–26

120 μa./cm.² show, for example, that only half the light output comes from the β process. If α and β processes are assumed to be of equal efficiency in light production from electronic energy, then the current density producing the constant of 160 is not 120, but 60 μa./cm.² This correction would considerably straighten the curve of the growth constant, making its rise nearly linear with

current density (see figure 6, *d*). In contrast to this, the decay constant changed by less than 20% over the whole current-density range of the diagram, emphasizing the importance of current density in growth processes as required by theory.

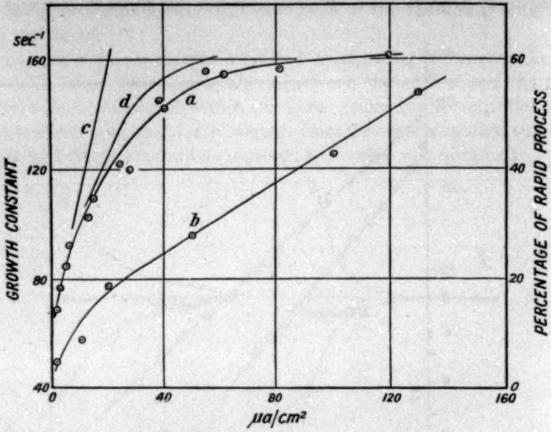

Figure 6. Change of growth processes for Zn_2SiO_4 . Mn with current density.

a. Slow growth constant observed.
b. Percentage of rapid process observed.
c. Original slope of $(a) = A'i + B$ from equation (4).
d. Corrected form of (*a*): see text.

If $B = 67$, $A' = 4{\cdot}3$ when i is measured in μa./cm².

In general the manganese class materials had a growth process well matched by a single exponential at low current densities. At increased current densities a good match could be made by two exponentials superimposed.

Effect of activator content

Interesting results were obtained by varying the amount of activator in zinc silicate. Above about 1%, manganese produces absorption bands in the visible region and a marked change of decay rate, shown in table 4.

Table 4. Change of decay constant of Zn_2SiO_4. Mn with Mn content

% Mn	Decay constant (sec.⁻¹)	Peak of emission (A.)	Relative efficiency
0·05	67	5245	8·4
0·2	67	5250	10·0
0·5	67	5255	9·7
1·0	80	5263	6·0
2·0	113	5291	5·5
5·0	550	5333	1·3

Very similar results for Zn_2SiO_4 . Mn were reported by Fonda (1939) using ultra-violet radiation.

Effect of temperature

With similar arrangements to those described in the measurements on the Ag class, the effect of reducing the temperature of the phosphor to $-180°$ c. was examined for Zn_2SiO_4 . Mn and ZnS . Mn. Only very slight differences in the decay constants were observed and they were too close to experimental error to be significant. Measurements on Zn_2SiO_4.Mn were also made at temperatures up to 250° c. : the growth constant seemed to be little affected by the change, but there was an increase of about 15% in the decay constant. As for the silver class, there is no such change as would be expected for processes involving ionization.

Survey of the more complex materials

An increase in the sensitivity of the apparatus made it possible to separate different parts of the emission spectrum in some cases and show that they had different decay constants.

In place of the photocell unit previously used, two new photoelectric multipliers were obtained, one with a maximum sensitivity in the red and the other in the blue-green. With these it was possible to examine materials hitherto too dull for test, and also, with the help of suitable filters, to separate different spectral regions.

Table 5

Material	Emission spectrum	Current density (μacm^{-2})	Decay constants (sec.^{-1})
Mg_2SiO_4 . Mn	Red band	2–500	17–23 ( ) (α), 10–37%
	Blue band	2–500	*c*. 3000
Al_2O_3 . Cr	Red lines	50–400	210–250 (β)
	Green band	50–400	*c*. 2500
$ZnAl_4O_7$. Cr	Red lines	5–400	60 (β) *c*. 1000 (α)
	Green band	5–400	1400–3000
$ZnAl_4O_7$. Co	Red and blue bands	5–400	200–800 1500–5000
$ZnAl_4O_7$. Sm	Orange lines	5–400	400 (β)
	Blue band	5–400	2000

The most interesting material was $ZnAl_4O_7$.Cr, which has an emission spectrum consisting of (1) a green band with a rapid exponential decay and a constant increasing from 1400 to 3000 sec.^{-1} with increasing current density; and (2) a group of red lines with an exponential constant of 60 sec.^{-1}. With increase of current density the proportion of the green component increased, and at the same time a rapid red process appeared with a constant of about 1000 sec.^{-1}. At $-190°$ c. the green also increased in relative proportion, the constant being apparently unchanged; the red was too weak to permit a determination. It appears that processes typical of both the Ag and Mn classes may occur in the same phosphor. It is not clear whether both types can be attributed to the one activator or whether the rapid decay regions, for example the higher frequency bands in $ZnAl_4O_7$. Cr and Mg_2SiO_4. Mn, are due to the base material. Zinc silicate, which when pure has a decay of the Ag type, and when activated by Mn joins the Mn class, might by careful control of the degree of activation yield a mixed phosphor of this type.

The different results are listed in table 5. They include a group of materials with $ZnAl_4O_7$ as a base and a number of different activators; this is of interest in expanding the Mn group of activators by Co, Cr, and Sm. The case of Sm needs further examination, for it will be seen that CaO.Sm appears among the results of the Ag class in table 1.

In the last two materials, separation of spectral regions was incomplete owing to low luminous efficiencies.

(c) *Processes lasting more than* 1 *second*

So far no direct evidence for ionization processes has appeared in these results. The lack of marked temperature dependence suggests that ionization plays only a subsidiary part. On the other hand it seems certain that the primary process responsible for the absorption of the energy of the incident electrons must be one of ionization. There is direct evidence of its occurrence in some phosphors, particularly under photo-excitation: temperature-dependent decays of a hyperbolic form have been measured, and photoconductive effects have been observed in cases where no such decays were apparent, for example in ZnSe (Miller and Strange, 1938); and they are also reported for materials such as Zn_2SiO_4.Mn (Randall and Wilkins, 1939). To round off the work already described, it was thought important to measure ionization processes in phosphors under electronic bombardment even if present in very small proportions only. In the work described it was not possible to make the comprehensive survey of the subject that was intended. All that was achieved was the development of a new method of measurement suitable for examining phosphorescent decays when present in amounts less than 1% of the total initial intensity, and the application of this method to a few materials over a limited range of current density.

The general lay-out of the apparatus was very similar to that given in figure 3 except that (1) the cathode-ray monitor tube had the usual two pairs of plates instead of four pairs; (2) in place of the matching unit a timing unit for calibrating the rate of scan of the monitor tube was included; (3) the time base was extended to several seconds in a single sweep across the monitor tube, and (4) the H.T. unit of the multiplier had high output and low impedance in order to avoid the effects of changes in voltage during the initial sudden drop of light intensity.

The device was operated as follows:—With appropriate current in the C.R. tube (adjust P 5), a value of R_1 and a suitable filter for the multiplier were selected to give a convenient deflexion on the monitor within the previously determined linear range. I_0 was thus recorded in relation to a base line adjusted to the same position by the bias control P 2 after each change of sensitivity of the multiplier. To record a decay curve starting from, say, 5% of I_0, the value of R_1 was increased by 20 times or the filter suitably changed. Depression of K caused the monitor to trace the decay curve on a y-scale 20 times as large as that for I_0, and still in the linear range. The decay curve was suitably disposed on the screen by adjustment of the scanning time, which was subsequently determined by calibration time pips recorded on a base line with the C.R. tube switched off. The y-axis was traced by alteration of P 2. The timing unit was free-running, hence the time signals on the x-axis could not be accurately synchronized with the depression of K, but the error was very small. The traces were read off from a scale on the monitor tube face, or photographed and measured later.

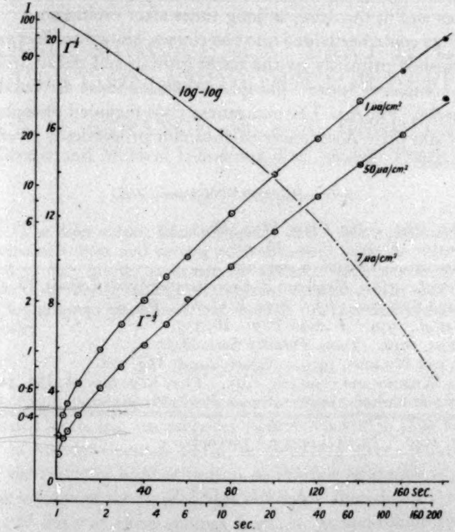

Figure 7. Decay of ZnS . Cu. Light units are arbitrary, but the same for the decays at 1 and 50 μa./cm² Logarithmic scales refer to decay at 7 μa./cm²

In most cases the traces for decay of ZnS . Cu and other materials of long afterglow gave a linear relation between time and $1/\sqrt{I}$ over the range $\frac{1}{2}$ to 4 seconds. It has since appeared that this period is too short to establish a bimolecular decay, as the extended results of figure 7 show. These decay curves for ZnS . Cu were taken at 3 kv. on an elaborated version of the apparatus, in which a cyclic scan showed successive portions of the same decay curve one below the other, continuing for any desired length of time. From figure 7 it seems doubtful if any part of the decay process can be accurately classed as bimolecular, and even if we assume a power law in which $m \neq \frac{1}{2}$ (equation (1)), the logarithmic plot in figure 7 shows no finality in the slope up to 3 minutes' decay. A more thorough investigation of this part of the problem is evidently required.

§4. CONCLUSIONS

The experimental work shows that processes of exponential form occur widely in growth and decay of light output from inorganic phosphors. The random nature of the processes is further emphasized by the small effects of temperature change; that is, the experimental processes are not bimolecular changes which have become pseudo-exponential by excess of one "reactant", as suggested by Fonda (1939) for part of the willemite decay process at low temperatures.

The simple theory fails to account for the complexity of the results, especially in the changing proportions and speeds of the simultaneous processes which were measured. There was no definite evidence in favour of bimolecular decay, though non-exponential processes were found to be present in the growth at low current densities and in the decay at long times after excitation.

The phosphors could be divided into two classes, known as *silver* and *manganese* classes, distinguished primarily by the faster growth and decay processes of the former, which contained "pure" phosphors besides those activated by Ag, Cu, Bi, Zn and, possibly, Pb, Au. The manganese class included phosphors activated by Mn, Sm, Cr and Co. A summary of phosphor properties is given in the third paper of this series.

REFERENCES

Becquerel, 1861. *Ann. Chim. Phys.*, **62,** 5.
Fonda, 1939. *J. Appl. Phys.*, **10,** 408.
de Groot, 1939. *Physica*, **6,** 275, 393.
Lenard *et al.*, 1928. *Handbuch d. Experimentalphysik, Leipzig*. **23,** 176.
Miller and Strange, 1938. *Proc. Phys. Soc.*, **50,** 374.
Nelson *et al.*, 1939. *J. Appl. Phys.*, **10,** 335.
Pringsheim, 1939. *Trans. Faraday Soc.*, **35,** 87.
Randall and Wilkins, 1939. *Nature, Lond.*, **143,** 978.
Randall, Wilkins and Garlick, 1945. *Proc. Roy. Soc.*, A, **184,** 347–433.
Schleede and Bartels, 1938. *Z. tech. Phys.*, **19,** 364.
Strange, 1938. *Trans. Faraday Soc.*, **35,** 95.
Strange, 1940. *Thesis* (London University).

APPENDIX

Description of figure 3

The apparatus includes—

(1) A key K which, when pressed, raises the potential of point A from 76 v. to 170 v., at the same time disconnecting the valve anode from the 360 v. supply.

(2) A scanning unit B consisting of a condenser charged through a MSP4 valve acting as a constant-current device. Its operation is controlled by the potential of point A which, through potentiometer P3, determines the potential of the grid of the valve. The rate of scan is controlled by P3.

(3) A cathode-ray tube C, which may be demountable, containing the screen under test. The cathode of the electron gun is connected to point A and the modulator to potentiometer P5, which, with the key raised, determines the current through the tube. Depression of the key makes the cathode 94 v. positive and cuts off the current.

(4) A photoelectric multiplier D with an output stage consisting of two KTZ41 valves in push-pull, giving a linear response of 100 v. either positive or negative in sense.

(5) A matching unit E consisting of condensers and resistances in series connected to the 360-v. supply and the key. With the key up, the voltage across the condensers is the difference between A and 360. Depression of the key causes the two sides of the condensers to be "shorted" through their respective resistances; the potential falls exponentially in each case.

(6) A special monitor tube F with four sets of electrostatic deflexion plates used for observing the results of mixing signals from the photoelectric unit and the matching unit.

CATHODO-LUMINESCENCE: PART II. CURRENT SATURATION AND VOLTAGE EFFECTS

By J. W. STRANGE and S. T. HENDERSON,
Electric and Musical Industries, Ltd., Hayes, Middlesex

MS. received 10 *September* 1945

ABSTRACT. The light output from phosphors under steady electron beams has been measured at constant voltage and varying current density, and thus "current saturation" has been found to vary greatly in extent for different materials. Similarly the change of light output with varying voltage at constant current density has shown different characteristics for different phosphors, but without the expected variations on changing the current density.

§ 1. INTRODUCTION

The preceding paper on cathodo-luminescence (referred to as Part I) was concerned with the variation of light output with time from phosphors under a steady electron beam, or decaying after excitation. We now consider the variations of light output at *equilibrium excitation* when the current density or the voltage of the impinging electrons is altered.

In the early days of experimental work on cathodo-luminescence, Lenard (1903) proposed for the light output from a phosphor the law

$$L = ki(V - V_0)^q, \qquad \ldots\ldots(7)$$

where k is a constant, i is the current, V is the voltage of the impinging electrons, V_0 is a "dead voltage" typical of the phosphor and attributed (Lenard and Saeland, 1909) to damaged surface layers, and $q = 1$. This law has been found to have limited application only. The fact that in general L increases less rapidly than i has long been known (as *current saturation*), while several workers have attempted to fit their results to expressions like equation (7) in which q has some value, not necessarily integral, between 1 and 3. A square law has frequently been used in this way (Brown, 1937; Nottingham, 1937, 1939).

These and other investigators have studied another property of phosphors, namely secondary emission characteristic, which by its variability adds to the difficulty of accurate light-output measurements. In general these materials show a ratio of less than one secondary electron to each primary one up to an incident energy of, say, 100 v. For greater energies, up to several thousand volts, the ratio is greater than one; then at a point commonly called the second cross-over the ratio falls to one, and for still higher beam voltages the phosphor tends to charge down towards cathode potential and maintain the voltage difference of the second cross-over. Surface contamination caused, for example, by tube baking, or by burning of the phosphor under excessive bombardment, are two causes of changes in the second cross-over point. Other experimental difficulties met in the accurate measurement of the light output at current densities above a few μa./cm.^{2} are:

(1) the raising of the temperature of the phosphor, with consequent change in efficiency;

(2) the burning of the phosphor, resulting in the destruction of its fluorescent properties, particularly under steady excitation at high current densities;

(3) the non-uniform distribution of current in the exciting electron beam produced by the usual electron-gun systems.

Most of the published results for the light output from fluorescent screens have been made using a scanning electron beam, i.e. moving at speeds of about 10^5 cm./sec. across the screen. It is very difficult to define accurately the conditions used in such an operation and to interpret the results obtained.

Results of measurements under steady excitation have been published by Nottingham (1937), who used an evacuated, sealed-off tube containing six phosphors deposited on glass plates fixed close to a central electrode, which could be used for measuring the current. The improved method described in his later paper (1939) still appeared to be open to doubt on the points raised above, especially regarding surface contamination which affected the second cross-over point.

§2. RELATION OF LIGHT OUTPUT TO CURRENT DENSITY AT CONSTANT VOLTAGE

The apparatus used consisted of the demountable tube shown in figure 8. With a rapid pumping system this worked satisfactorily at low current densities, though at high current densities the difficulties listed above were encountered, as in the case of sealed-off tubes. However, by the use of a new ballistic method these difficulties were successfully avoided.

A thermionic cathode of the impact type, (Ehrenberg, 1939), in which a tantalum cap is heated to about 2000° c. by electronic bombardment, yielded the necessary high currents. The beam, rendered roughly parallel by the magnetic coil N, was defined by the electrodes A, B and C, of which A prevented secondary electrons moving from B to C, B defined the beam falling on the sample and was kept about 9 v. negative to A and C, while C collected secondary electrons from the sample D for beam-current measurement. A and C were at the final anode potential, B had to be in good contact with the tube wall, leaving no gaps, but separated from the graphite coating E as shown.

The glass plate carrying the eight sample phosphor screens was mounted on a ground joint in the glass end-plate. The samples could thus be brought into the same position under the electron beam in turn. The end-plate was sealed on with Apiezon compound Q.

Preliminary tests were made to confirm the accuracy of beam-current measurement by means of secondary emission to electrode C; also to check the current distribution across the electron beam, which varied by less than 10% at low current densities, but became less uniform at increased currents, with the effect of spreading out changes in slope of the (light *vs.* current density) curve.

The beam-current measurement was made with a mirror galvanometer F with universal shunt, giving sufficient sensitivity for steady or ballistic measurements· The light measurement employed a similar galvanometer G with a shunt, receiving the output from a photo-cell H and bridge-type D.C. amplifier K. This amplifier used two LP2 triodes operating over the lower part of their characteristics, and gave a very steady and linear output when properly balanced. Colour or neutral filters L were used as required.

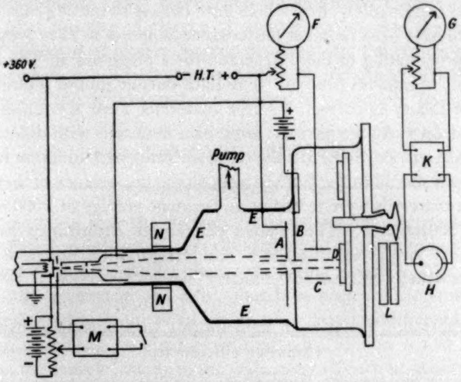

Figure 8. Apparatus for light-output measurements.
Electron paths shown by broken lines.

Ballistic measurements were made by applying a potential pulse to the cathode modulator system of the electron gun for a predetermined time, by means of the multivibrator M, and measuring simultaneously the photo-cell output and the beam current, each galvanometer being critically damped instead of freely oscillating. Control experiments showed this method to be reliable (cf. Masius, 1925). The pulse length in ballistic measurements was fixed at 0·1 second, which reduced the error due to finite growth period of the light output to about 2% for willemite and less for other phosphors. The upper limit of the pulse was determined by the possibility of overheating or burning the samples. The absence of serious effects was checked by using a screen of $CaWO_4$, of which the light output is very sensitive to temperature changes.

Experimental results

Measurements made on a wide variety of materials showed that in all cases there was a linear relation between light output and current at very low current densities, as indicated by equation (5) (Part I). This fact was employed in all subsequent measurements to eliminate various factors irrelevant to the main purpose of the tests. Among these may be mentioned the variation in the intrinsic brightness and in the position of the emission band, differences in thickness of the screen and in the light absorption in the materials. With the original linear slope as a basis for comparison, and taking it as unity in all cases, such variables could be eliminated. The efficiencies given will thus be relative to the efficiency of the initial linear range. It may be noted that the absolute efficiency reported for the most efficient materials is of the order of 10% (Leverenz and Seitz, 1939).

Three materials were selected for detailed examination: ZnS . Ag which is typical of the group of rapidly decaying phosphors called the *Ag class* in Part I, and ZnS . Mn and Zn_2SiO_4 . Mn, which both belong to the *Mn class* of phosphors. These last two have been found at low current densities to have growth and decay processes approximating to those expected for a phosphor in which the luminescence is a simple random process. The light output should follow equation (4), Part I, if the theory is correct. These materials were also chosen because in ZnS . Mn and ZnS . Ag we have the same base materials with different activators and in ZnS. Mn and Zn_2SiO_4. Mn the same activator with different base materials. The comparison should throw further light on the importance of these differences.

The three materials were tested at an incident voltage of 5000 over a current range up to 300 μa./cm.^{2} The relative efficiencies determined by the method described above are given in table 6; these figures do not imply that all the materials were equally efficient at the lowest current densities.

Table 6. Variation of light output with current density
(Relative efficiencies)

Current density (μa./cm.^{2})	1	50	100	150	200	250	300
ZnS . Ag	100	100	100	100	100	97	93
ZnS . Mn	100	86	72	62·5	57	53·5	50
Zn_2SiO_4 . Mn	100	68	50	40·5	36	32	30
„ calculated	94	25	14	10·5	8	6	5

The calculated values for Zn_2SiO_4 . Mn given in table 6 were derived from the simple theoretical expression for light output and the values of A' and B previously found (Part I, equation (5) and figure 6). The failure of these calculated values to agree even approximately with experiment was to be expected from the disagreement already shown in figure 6, where the growth constant failed to conform to the linear variation with current density which the theory predicted, and this failure persisted even after correction for the proportion of rapid process (see Part III).

The efficiency figures in table 6 seem to indicate fundamental differences in properties between the phosphors. The influence of methods of preparation was investigated, and here only minor changes were observed. Zn_2SiO_4 . Mn made by four different methods gave values of 31, 34, 41 and 44 for the relative efficiencies at 250 μa./cm². A similar group of ZnS . Ag materials gave 90, 93, 95 and 97.

On the other hand, marked changes were observed when the activator content was changed in Zn_2SiO_4 beyond the range of about 0·2 to 1·0% already remarked upon in the decay measurements (Part I). Using the same set of Zn_2SiO_4 . Mn materials with manganese varying from 0·05 to 5·0%, light-output curves were obtained in which, over the range 0·2 to 2·0% Mn, there was a progressive increase in the length of the initial linear part. Above and below these proportions the change in form was rapid, with the slope observed at 250 μa./cm² lowest for 0·05% Mn, about equal for 0·2 to 2% Mn, and still greater for 5% Mn.

Measurements on the three principal materials were made up to current densities of 1 ma./cm², but considerable difficulty was experienced in ensuring reproducible conditions. The results can be summarized thus for ZnS. Mn and Zn_2SiO_4 . Mn: the rate of change of slope observed over the current range 20–200 μa./cm² lessened at higher current densities, and the curves tended again towards a linear relation. There appeared to be no indication of an approach to a maximum value of the light output as suggested by the simple theory.

§ 3. VOLTAGE EFFECTS

(a) *Growth and decay processes*

It seemed possible that the probability constant for excitation, A (or A'), introduced in Part I, might vary with the density of exciting radiation, and thus with the voltage of the impinging electrons. Using Zn_2SiO_4 . Mn, experiments were made in the apparatus of § 3 (b), Part I, to measure the growth constant, which the simple theory requires to be $Ai+B$. Some additional difficulty was experienced in obtaining satisfactory current distribution across the electron beam falling on the sample, owing to the wide range of current density and voltage used. Thus the results will not show the maximum changes that may in fact occur. Table 7 A shows the usual increase of growth constant with current density, but no appreciable effect by change of voltage; table 7 B shows that the initial rapid growth process decreases in amount with voltage increase at any current density (except the lowest used). Table 8 gives comparable results for the proportion of rapid decay process, α. Other materials behaved similarly, for example $ZnAl_4O_7$. Mn and $Zn_2B_2O_5$. Mn.

Some connexion may be seen between some of the changes in growth and decay process so far recorded. In Part I it appeared that increased current density at constant voltage caused increases in proportion of rapid growth and decay processes. In the present case we have varying voltage, and, since faster electrons penetrate more deeply, the average energy dissipation per second per unit volume of phosphor is greater for smaller incident voltage at constant power input. If we consider both cases in terms of power input it is seen that in each an increased amount of rapid process accompanies increased energy dissipation per second per unit volume. Compare for example the groups of figures marked in tables 7 B and 8: the percentages are greater for lower voltage at equal power input. Current

Table 7. Growth process for Zn_2SiO_4 . Mn at different voltages and current densities

A. Constants for main process (sec.^{-1})

Current density (μa./cm.^{2})	Voltage (kv.)						
	1	2	3	4	5	6	8
1	72	78	79	79	81	73	70
10	110	111	114	118	118	108	111
20	121	132	130	122	125	122	122
50	137	147	139	132	137	143	143
100	139	139	139	135	134	133	137

B. Percentage of rapid process

Current density (μa./cm.^{2})	Voltage (kv.)						
	1	2	3	4	5	6	8
1	9	15	11	9	6	15	22
10	25	17	14	9	3	5	4
20	33	20	17	26	17*	9	10
50	44	33*	35	38	25	24†	25
100	56*	46	47†	46	38	38	38

Note: Four independent groups of figures have been selected: those marked * in table 7 B (0·1 watt/cm.^{2}), those marked † in table 7 B (0·3 watt/cm.^{2}), and similar groups in table 8 (* 0·2, and † 0·8 watt/cm.^{2}).

Table 8. Percentage of rapid decay process (α) for Zn_2SiO_4 . Mn at different voltages and current densities

Current density (μa./cm.^{2})	Voltage (kv.)						
	1	2	3	4	6	8	10
0·3	10	8	10	5	5	5	2
2·5	8	10	6	6	2	0	0
20	20	12	10	8	6	8	10*
50	25	20	17	12*	12	15	15
100	25	16*	18	17	14	18†	18
200	—	—	27	26†	22	22	20

saturation may be similarly treated in terms of energy, and it is shown in Part III of this series that a greater percentage loss of efficiency occurs for the lower voltage if the power input is kept constant.

(b) *Value of voltage exponent q*

If, as is generally assumed, the rate of energy dissipation in depth varies with the speed of the electrons at the particular depth considered, it might be expected that under different current-saturation conditions different values of q (equation (7)) would be found. A further series of experiments was directed towards the determination of q at various current densities. The apparatus was similar to that described in §2. Certain modifications were called for by the need to maintain similar conditions of current distribution at widely varied voltages, whereas previously the voltage was kept constant (except in the work described in §3 (a)). Unless this uniformity is maintained, the variations in light output due to current density changes may easily obscure the effect of voltage changes. It was also desirable to increase the current density as much as possible above the previous maximum of 300 μa./cm^2, to emphasize the expected changes of q with current density.

The electron gun consisted of a cathode of the impact type (Ehrenberg, 1939), a modulator, an accelerator, and an anode containing an aperture of 0·1 inch diameter. The modulator was used only for applying the controlling pulse and was at cathode potential during the period of excitation. The current through the tube was then dependent on the temperature and emissivity of the cathode and the potential of the accelerator, with only slight dependence on the anode voltage. No focusing of the electron beam was attempted, as this would obviously alter with anode voltage; in consequence the size of the beam striking the screen was also independent of the voltage and was determined only by the size of the aperture in the anode cylinder and its distance from cathode and screen.

The usual method of operation was to vary the current in the tube by altering the cathode temperature. The latter was difficult to control with precision, and it was necessary to take a number of readings scattered about the required current density value and construct from them a small section of the current density *vs.* light-output curve at that voltage. The value of the light output at the required current could then be read off with a high degree of accuracy. The same ballistic method, with pulses of 0·1 second, was used as in the experiments of § 2.

Experimental results

Some of the results for the three main materials, $Zn_2SiO_4 . Mn$, $ZnS . Mn$ and $ZnS . Ag$ are given in figures 9 to 11. The most striking of these are the results for the first material. Over the wide range of voltage from 300 to 9000 v., and without any adjustment for V_0, there was no variation in the value of q, and over a considerable change of current density, from 5 to 550 μa./cm^2, the variation was very small.

The other two materials needed corrections for V_0 to straighten the tail of the curve. With the extension of the measurements to voltages close to these values of V_0, the corrections could be made with certainty. The low values of V_0 are evidence that the screens suffered a minimum of contamination. $ZnS . Mn$ showed some instability at high currents and voltages, and a low second cross-over of the secondary emission curve, but otherwise both sulphides behaved like $Zn_2SiO_4 . Mn$ over a wide voltage range, with a similar small, if not negligible, change of q with current density.

Besides the failure of these materials to show any marked change of q with current density, there appeared quite definitely the fact that q is characteristic of the phosphor and not the same for all materials. The possibility that this variation was due to the absorption of light in the phosphor was considered. It was,

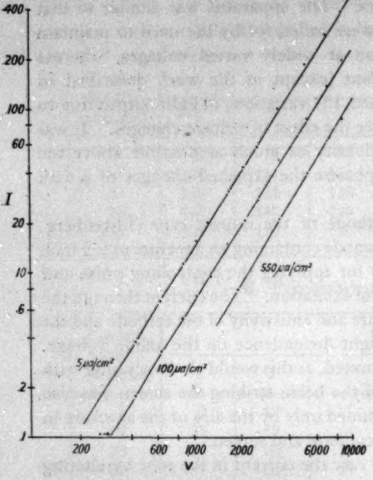

Figure 9. Voltage-brightness relation for $Zn_2SiO_4.Mn$. Light units different for different current densities.

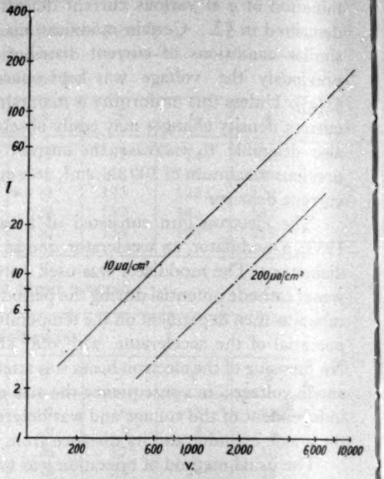

Figure 10. Voltage-brightness relation for ZnS . Mn. Light units different for different current densities.

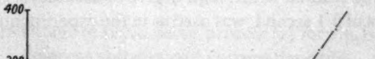

Figure 11. Voltage-brightness relation for ZnS . Ag. Light units different for different current densities.

however, ruled out by measuring the light emitted from the back (i.e. the side facing the electron source), and from the front of the screen for ZnS . Ag. There was no difference between the slopes derived from the two sets of figures. Some such effect would be expected, perhaps only in a much higher voltage range.

The results are collected in table 9.

Table 9. Values of q and V_0

Current density (μa./cm^2)		5	10 front	10 back	100	200	550
Zn_2SiO_4 . Mn	q	1·53			1·54		1·59
	V_0	0			0		0
ZnS . Mn	q		1·0			0·98	
	V_0		200 v.			250 v.	
ZnS . Ag	q		1·36	1·36			1·32
	V_0		150 v.	150 v.			150 v.

The theoretical aspects of the work recorded in this paper, and in the previous one on growth and decay processes, are considered in the third paper of this series.

ACKNOWLEDGMENTS (PARTS I AND II)

Our thanks are due to Mr. I. Shoenberg, of E.M.I., Ltd., for permission to publish this work; also to Mr. G. E. Condliffe and our colleagues of the Research Laboratories of E.M.I., Ltd., and especially to Messrs. L. F. Broadway, E. W. Bull and H. Hirsch. The experimental work was done by J. W. S. during his service with the company up to 1941, apart from the preparation of phosphors and spectrographic work by S. T. H., who also prepared the paper for publication.

REFERENCES

Brown, 1937. *J. Opt. Soc. Amer.*, **27,** 186.
Ehrenberg, 1939. British Patent 523,240.
Lenard, 1903. *Ann. Phys, Lpz.*, **12,** 449.
Lenard and Saeland, 1909. *Ibid.*, **28,** 485.
Leverenz and Seitz, 1939. *J. Appl. Phys.*, **10,** 479.
Masius, 1925. *Phys. Rev.*, **25,** 211.
Nottingham, 1937. *J. Appl. Phys.*, **8,** 762 ; 1939. *Ibid.*, **10,** 73.

CATHODO-LUMINESCENCE: PART III. DISCUSSION OF RESULTS

By J. W. STRANGE and S. T. HENDERSON,
Electric and Musical Industries, Ltd., Hayes, Middlesex

MS. received 22 *December* 1945.

ABSTRACT. The interpretation of the experimental results in the preceding two papers is attempted, but cannot be adequate without more knowledge about electron absorption in phosphors. Some possibilities are outlined with regard to the mechanism of luminescence. There is some evidence in favour of a new type of voltage absorption law.

---

§1. INTRODUCTION

The two earlier papers in this series have given the experimental facts of the present study, and this paper will examine theoretical matters, commencing with qualitative aspects of luminescence and proceeding to a more quantitative discussion of some of the data. We may first enquire what bearing the results have on current views of the mechanism of luminescence, considering mainly the evidence from the growth and decay processes. This evidence has already been shown to be too complicated to fit the simple theory of the first paper (referred to here as Part I), and there is also no clear agreement between the theory and the observations on light emission at equilibrium excitation. A more thorough knowledge of the processes of electron absorption in crystals is required before much progress can be made in interpreting the results: this need will appear again later when empirical quantitative methods are introduced to deal with the voltage results. Recent developments in the theory of trapping states in phosphors may provide a valuable new approach to some of the unsolved problems. So far, however, these ideas of Randall and his co-workers (1945) seem to be applicable mostly to decays of a much longer order, produced by photo-excitation at energy densities of much lower value, than those used in the present work.

§2. QUALITATIVE ASPECTS OF LUMINESCENCE

The distinctions between excitation by ultra-violet radiation and by an electron beam are well known. Photo-luminescence depends on the presence of absorption bands which may be of different kinds in the same phosphor, giving different decay rates (Kröger, 1939), and possibly different luminous efficiencies. In cathodo-luminescence all atoms of the phosphor within the range of incident electrons must take part in the primary absorption of energy. This may cause excitation without ionization of (*a*) the atoms of the base material, or (*b*) the impurity "activator" atoms; or it may produce ionization of both kinds of atom. Ionization is generally agreed to be the more probable result, and where the activator content is small, as in the *Ag class*, far more atoms of the base material than of the activator must undergo this process. Since luminous efficiencies are of the order of 10% in cathodo-luminescence, and are similar for materials of widely differing activator contents (0·01% in ZnS . Ag, 1% in Zn_2SiO_4 . Mn), and since the active centre determines the frequency of the emission band or lines, it is a necessary conclusion that energy absorbed from the electron beam by the base material is transferred to the activator atoms before luminescence occurs. This transfer may be effected in two ways: either by movement of electrons in the conduction band, and at the same time of positive holes in levels below the ground state in the active centres, or, on the other hand, by an " exciton " mechanism (Mott and Gurney, 1940, p. 208). The first alternative would be supported by the presence of photo-conductivity, which is found in some, but not all, phosphors. The second is in agreement with the small amount of very slow phosphorescence observed in most cases, for it implies that the final excitation of the centre will not be an ionization, and therefore exponential forms may be expected for growth and decay processes.

The lack of temperature dependence in growth and decay (Part I) is decidedly in favour of the hypothesis that the light-emission processes do not directly involve ionized electrons, but this cannot be taken as a conclusive proof. From the converse point of view, well defined non-ionizing processes may sometimes show a type of temperature dependence resembling the " freezing in " of luminescence in materials where ionization precedes emission. For example, in $ZnAl_4O_7$. Cr the transition of a 3*d* electron in Cr^{+++} gives a doublet ^4F—^2G near 6870 A. which has satellite lines due to intramolecular Stark effect. The lines of greater frequency than the doublet are much reduced in relative intensity by fall of temperature, and almost disappear at −180° c. The lines of smaller frequency than the doublet are not greatly affected in intensity, relative to the main lines, by temperature change between −180° c. and 300° c. It would be of interest, though difficult, to determine growth and decay constants and luminous efficiency for separate lines in this emission spectrum. Again, pseudo-exponential processes may be temperature dependent (Part I, § 4).

The evidence of growth and decay measurements suggests the presence of more than one activator level (or group of levels), with different characteristic speeds, their relative participation in the process of luminescence varying with current density. Shift of emission-band position with altered density of excitation radiation would support this view, though conversely the absence of spectral shift does not rule out the possibility of more than one level. Certain small effects have been found on changing from electron excitation to the less intense irradiation by ultra-violet (Henderson, 1939), but an increase of current density by 40 times in an electron beam at 5 kv. has been found to produce no change in the peak position of the ZnS . Ag band, nor has the increase of voltage from 2 to 10 kv. at constant current density significantly altered the position of the ZnS . Zn band.

With regard to emission bands, it is probable that some small amount of energy transfer occurs by optical absorption. When a phosphor has two emission bands, one may be of sufficiently high frequency to excite the other in a separate screen, or in layers of the same screen further from the exciting radiation, for example in ZnS . Cu or ZnS . CdS . CuAg. The absorption bands concerned are not usually intense enough to make the effect important.

The assumption of two simultaneous emission processes of exponential form reduces to some extent the disparity between the simple theory of Part I and the experimental results. The calculated brightness figures in table 6, Part II, are for the slow β process: the difference from the experimental values may then be taken to represent the contribution of the rapid α process, if both α and β are equally efficient in light production. Hence by calculation the percentage of rapid process increases from 6% at 1 μa./cm.^{2} to 63% at 50 μa./cm.^{2} and 78% at 200 μa./cm.^{2}. These percentages are increased if the non-linear values of the growth constant are taken from figure 6, Part I. The observed proportions of rapid growth and decay processes were, however, far less than these calculated values. Also, inconsistent values of A' were found by calculation from the experimental data for light output, growth, and decay respectively. This modification of the simple theory is, therefore, of no great use in reconciling theory with experiment, at any rate for Zn_2SiO_4 . Mn.

The effect of collisions of the second kind has been considered as a possible mechanism to account for increasing growth and decay rates with increasing current density. If the collision of an electron and an excited centre were to cause an immediate transition with luminescence, a kind of acceleration effect would appear as the electron density increased. The most noticeable result would be a nearly instantaneous fall of light output at the end of the excitation period owing to the cessation of this acceleration. The fall would increase with current density.

There is in fact a very rapid fall at high current densities under scanning conditions. Screens of Zn_2SiO_4 . Mn and ZnS . Mn were scanned by a spot of high current density ($\simeq$5 amp./cm.^{2}), when some proportion of the decay occurred in a process lasting not more than a few microseconds. This may be regarded merely as the increased amount of rapid decay (α process, see table 2, Part I); alternatively it could be attributed to an acceleration effect, but for the fact that there is a similar rapid rise at the start of excitation by high current densities, which would not be expected as a consequence of the effect in question. The experiments of Part I showed the growth and decay curves to be always similar in form if their ordinates (light output) were measured from the initial value, zero or I_0 respectively. There was no suggestion of the dissimilarity which would arise from an acceleration process. A more plausible explanation of the very fast decay at high intensities has been given by Mott and Gurney (1940, p. 217), who believe that the lifetime of the excited states determines the decay rate under these conditions.

A more detailed mechanism is required to explain the changes of speed with current density in the separate α and β processes. If it is assumed that the lifetime of excited metastable states in the activator atoms may be of the same order as the observed decay times (cf. Mott and Gurney, p. 209), the following scheme seems to be worth consideration.

The main excitation occurs, presumably, by ionization of atoms of the base material, followed by the recombination of ions with electrons and the generation of exciton waves. The excitons raise electrons of the activator atoms from the ground state X to non-ionized metastable levels Y. X and Y are located, according to the current theory of luminescence, in a forbidden energy zone between the highest full level F of the matrix atoms and the normally empty conduction band. Luminescence arises from the transition Y→X, which gives exponential forms largely independent of temperature changes. The separation of X and Y must, in effect, have a Gaussian error distribution to account for emission band width and shape, though this variation must not affect the transition probabilities.

Assume now that the X-Y group of levels may take up varying positions in the forbidden region, which may happen for activator atoms differently situated in the lattice. It seems likely that the probabilities of excitation X→Y and emission Y→X will vary according to these positions. That is, the energy difference F – X will affect the probabilities. Constancy of spectral emission will be assured if the separation X – Y remains the same for any value of F – X, apart from the differences causing band width. There are thus available a number of transitions of different effective speeds. For the experimental conditions of excitation over relatively long periods of time, this should result in increased rates of growth and decay with increased density of excitation, owing to saturation of the more slowly decaying states.

This preferential saturation should not be observed for times of excitation short compared with the decay times, unless another condition is imposed. Experiments quoted above showed that in scanning, with excitation times of about a microsecond for each screen element, increases occurred in decay rates similar to those under saturation excitation. The extra condition necessary to account for this is that increased probability of excitation must be accompanied by decreased probability of emission between the levels X and Y. The slower transitions Y→X will then be favoured at low excitation densities for any length of pulse. It is not evident whether the changes in probability would occur, in the sense mentioned, by increase or by decrease of the energy separation F – X.

The mechanism described could provide the variations with current density of a single process, e.g. the β process in ZnS . Ag. It would appear to need duplication at a higher range of excitation and emission probabilities to explain the simultaneous α process, and it is not certain if metastable transitions within the activator atoms of the *Mn class* can be entirely responsible for the slow exponential β process observed in cases where the decay constant is of the order of 10^2 sec.$^{-1}$.

Non-exponential processes, mainly in long-term decay, must have their rates determined by the behaviour of ionized electrons and the presence of traps, but this type of luminescence, though it has practical applications, is of minor importance in electronic excitation.

§3. DECAY RATES AND CURRENT SATURATION

This aspect of the experimental results can be considered without reference to the mechanism of electron absorption (see §4). Increase of energy absorption per second per unit volume of phosphor leads to increased overall rates of growth and decay. This appears most simply in the results for constant incident voltage (Part I). Similarly, power increase at constant incident voltage causes loss of light efficiency, but there is no clear relation between this change and the effects on growth and decay. The two kinds of effect may be independent and not occur necessarily over the same range of power input. Thus in the case of ZnS . Ag, a linear (light output *vs.* current density) curve between 0 and 200 μa./cm^2 at 5 kv. is accompanied by great increases in speed of growth and decay, while Zn_2SiO_4 . Mn exhibits much smaller changes in speed but a considerable fall in efficiency. Alternatively, the increased speed in ZnS . Ag may be the reason for the unaltered efficiency, simply by making the centres more rapidly available for re-excitation. This idea is supported qualitatively by considering phosphors in the order ZnS . Ag, ZnS . Mn, Zn_2SiO_4 . Mn. This is the order of increasing degree of current saturation, and also of decreasing speeds. The experimental values of A' at $i \simeq 0$ were 33, 15, and 4·3 sec.$^{-1}\mu$a.$^{-1}$ respectively; of B, approximately 10^4 to 10^5, 625, and 67 sec.$^{-1}$ respectively. These figures suggest that the comparatively large value of B is one reason for the lack of saturation effects in ZnS . Ag in the current range of the experiments.

§ 4. QUANTITATIVE ASPECTS

The processes of electron absorption and light emission in phosphors depend on several factors which might be measurable, but on which we have insufficient information at present:

(*a*) the manner in which electron velocities (or voltages) decrease with the thickness of matter traversed;

(*b*) the effective electron current, which may vary through the slab;

(*c*) the relation between absorbed energy and light output for any volume element of the phosphor, which involves the method of transmission of energy through the crystal as well as the efficiency of conversion to light at the active centres.

Apart from these major factors there are:

(*d*) scattering and straggling in the electron beam, which must be of some importance but will not be discussed; and

(*e*) light absorption and scattering, which occur in phosphor layers, but seem to have a negligible effect on the results (see Part II, table 9).

(a) *Current saturation*

In Part II, § 3 (a), some correlation was obtained by use of the energy parameter instead of separate V and i values. The current-saturation data for Zn_2SiO_4 . Mn may be examined in a similar way. Owing to the method of plotting current density *vs.* light output on relative efficiency scales, no values were available for direct comparison of the curves for different voltages. When plotted with the same initial slope, the 4 kv. and 10 kv. curves were very similar in shape over the range 0 to 600 μa./cm.^{2} The 4kv. curve is plotted in figure 12, *a*. The 10 kv. curve can be calculated from this, using the results of table 9, Part II, since brightness at 10 kv. = (brightness at 4 kv.) × $(10/4)^q$. The values of q for Zn_2SiO_4 . Mn in table 9 give a linear relation between current density and $(10/4)^q$: from this relation the appropriate values at each current density were taken to derive curve *b*, figure 12. If the experimental curve for 10 kv. is fitted to this theoretical curve, starting from point A, the other marked points are obtained, which show a satisfactory agreement to exist between the data of two different types of experiment. Next, *a* and *b* are replotted on abscissae representing incident energy per second instead of current density, which gives curves *c* (4 kv.) and *d* (10 kv.).

The increased saturation at lower voltage is now evident. There is a nearly constant ratio of 2·67 between the values of energy input rate at any relative efficiency in the range considered. Thus the power input required to cause a given loss of luminous efficiency is nearly proportional to the voltage used.

So far no assumption has been made about the distribution of the energy through the slab of phosphor. If the electron current is considered to remain constant during the absorption process, then it must be concluded that the voltage change in absorption is of the type mentioned in Part II, §3 (b), by which a given depth of matter produces a greater relative loss of speed in slow electrons than in fast ones. The Thomson-Whiddington law is the classical formulation of this process for fast electron absorption in gases and metal foils. It may be given as

$$y = a(V_1^2 - V^2), \qquad \ldots\ldots(8)$$

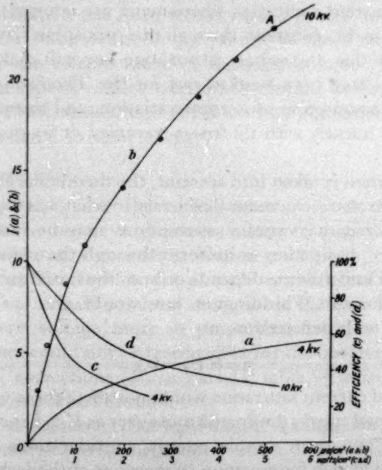

Figure 12. Current saturation in $Zn_2SiO_4 . Mn$.
Light units are arbitrary (curves *a* and *b*). Efficiency scale refers to curves *c* and *d*.

where y is the depth of penetration at which electrons of original voltage V_1 have been retarded to (a most probable) voltage V, and a is a constant for the absorbing material. In the comparatively slow speed range used in the present work, the value of a may decrease with V_1 (Allen, 1932), but apart from this correction it is not certain that equation (8) is applicable at all to the case of absorption in phosphor crystals. The discussion of decay processes in terms of power input, in Part II, §3 (a), is valid for any voltage absorption law since the deeper penetration of faster electrons distributes the incident energy over a greater volume of phosphor, and the decay rates change appropriately. In the next section it will appear that the Thomson-Whiddington law, or a similar one with indices greater than unity, leads to difficulties in interpretation.

(b) *Voltage exponent q*

The light output agrees experimentally with the expression $k(V_1 - V_0)^q$, where k, V_0, and q are constants. Explanations are required for:

(1) the constancy of q for a given material over a wide range of current density, showing no significant dependance on current saturation effects;

(2) values of q which differ from unity;

(3) differing values of q for similar materials.

These facts are more suitable for quantitative treatment than the current-saturation data. As before, the assumptions made will be rather speculative, in the absence of precise knowledge on electron absorption.

Current saturation and q

Provided that current saturation phenomena are ignored, and if the current density i is taken to be constant through the phosphor layer, simple energy considerations show that any voltage absorption law will make q equal to unity. Various other cases have been worked out for the Thomson-Whiddington law, with the additional assumption of current variation, and in each case q was unity whether i decreased linearly with thickness traversed or by power laws analogous to equation (8).

If current saturation is taken into account, the theoretical effect is to increase q markedly at higher excitation densities, contrary to what was found by experiment. To overcome this difficulty various assumptions may be made. For instance, suppose that energy dissipation is uniform through the phosphor, and that the power expended per unit volume depends only on the initial current i_0. For these conditions, the Thomson-Whiddington law would give $q=2$, while a linear decrease of voltage with penetration, or

$$y = b(V_1 - V) \qquad \ldots\ldots(9)$$

would give $q=1$, and current saturation would not affect these values. In general, to obtain $q=n$, the total energy dissipated must vary as $V_1{}^n$, hence the case of $q=1$ is the only probable one from simple energy considerations. This cannot be adduced as strong evidence for equation (9) because of the lack of support for the hypothesis of uniform energy dissipation.

A different assumption regarding energy dissipation, less rigid than the one just discussed, leads to the same independence of current saturation, and incidentally supports equation (9) as against equation (8). Consider a layer of phosphor of thickness y_1 in which electrons of the original voltage are just absorbed. Let dE/dy be the rate of dissipation of energy at a depth y in this layer. Now

$$\frac{dE}{dy} = i \cdot \frac{dV}{dy} + V \cdot \frac{di}{dy},$$

and from the particular modes of variation of i and V chosen, a curve may be drawn for the relation

$$\frac{dE}{dy} \text{ vs} . y.$$

Its area between $y=0$ and $y=y_1$ gives the total energy absorbed. Variations of

dE/dy with y mean that current-saturation effects may occur, but these need not alter the value of q if the overall degree of current saturation is the same for all incident voltages. This requires that for any incident voltage the same fraction of the total energy shall be dissipated at any given value of dE/dy.

A simple example is the linear decrease of both V and i with y, both V and i becoming zero together. Then dE/dy decreases linearly to zero with distance penetrated in the layer, and $q=1$. Other variations of i with y or V give the same result, provided the linear variation of V remains, as in equation (9).

If, on the other hand, the Thomson-Whiddington law is accepted, the elimination of differential saturation effects is not obtained with any of the proposed simple variations of i.

This brief discussion shows that the observed constancy of q under current-saturation conditions can be accounted for, provided that equation (9) holds, while no such simple explanation has been found with a power law like equation (8). This appears to throw some doubt on the validity of the Thomson-Whiddington law, at any rate for the voltage range of the experiments. Some independent support for the linear type of voltage variation is afforded by Fano's calculations on electron scattering (1940), which require the average penetration of slow electrons to be nearly proportional to their energy.

Voltage exponent q not unity

So far the only reasonable theoretical value of q not affected by current saturation appears to be unity; in practice, values of q greater than unity were obtained under conditions of current saturation.

A non-florescent surface layer on the phosphor, whether of foreign matter or of decomposed phosphor, would produce an apparent increase in the value of q by reducing the incident voltage more in proportion for slow electrons than for fast ones. This effect could not be responsible for the observed values of q, since the "dead voltage" corrections (V_0 in Part II, table 9) were much too small to account for changes in q of the order of 0·5.

Hitherto it has been assumed that light output is a function of dE/dy alone. It is doubtful if this position can be maintained in face of the results in question. It is likely that the apparent abnormalities arise from subsequent stages of the luminescence, following the primary absorption of energy, and that this cannot be treated by the methods outlined above.

From this point of view Fano (1940) has attempted a theoretical explanation of the work of Brown (1937), who found $q=2$ for $Zn_2SiO_4 . Mn$ between 200 and 1300 v. at low current densities. Fano examined electron scattering and diffusion, the resulting average depth at which excitons (or "wandering excitations") were produced for any voltage, and the probability of non-luminescent transfers of exciton energy at crystal surfaces. This is a much more complex treatment than the one we have given, and it is doubtful if it can be expected to give accurate q values, especially as many assumptions and approximations are made. The value of Fano's work lies not so much in its support of Brown's evidence, which is open to some criticism on the grounds of probable screen contamination, but in showing a method of approach, and indicating the possibility of values of q other than unity.

Different q values for similar phosphors

Events in the luminescent process subsequent to the primary absorption must also be responsible for the differences observed between the values of q for ZnS . Ag and ZnS . Mn. There is a difference in crystal composition between these phosphors, apart from the activators, the content of blende being probably from 20 to 40% in the ZnS . Ag, with the remainder wurtzite, while the ZnS . Mn contained 95% or more wurtzite. Differences of exciton diffusion and absorption by the active centres, as well as differences in surface properties, might well occur here. It would be of interest to measure q for a range of ZnS . Ag samples in which differing heat treatment had developed varying proportions of the two crystal forms.

To summarize this section on the voltage exponent: the absorption law $y=b(V_1-V)$ will account for the observed lack of dependence of q on current saturation, but no simple explanation can be made on the basis of equation (8). The processes by which the energy of the primary electrons is conveyed to the luminescent centres must be held responsible for the fact that q is not unity in some cases, while differences of crystal structure are the probable cause of the characteristic values observed. Current variations through the phosphor, which may have a profound effect on the results, have so far not been determined, and the experimental difficulties of such measurements would obviously be very great.

The experimental conditions may not have given the simplest results, and it would perhaps be better to use electron beams of much higher velocity, and very thin screens which only slightly retard the beams, together with means of avoiding secondary emission trouble.

§ 5. SUMMARY AND CONCLUSION

The similarities existing among all the phosphors tested are more remarkable than their differences. Those with Ag, Cu, Zn, Bi (? Pb, Au) as activators, and the "pure" phosphors, exhibit two exponential processes in growth and decay of luminescence, with comparatively high constants. The relative proportions of the processes change rapidly with current density, while the values of the "constants" change to a smaller degree. The activators all have atoms with completed electronic shells. Their concentration in phosphors should be small, of the order of 0·01%. The emission bands in this *Ag class* are wide and structureless and little affected by temperature; in general the wider the band, the faster the growth and decay processes, and the more rapid the decrease of light output efficiency with temperature increase.

Phosphors in the *Mn class* have as activators Mn, Cr, Co, with incomplete inner electronic shells. The exponential growth and decay processes vary slowly in speed and relative proportions with current density, and the constants are much lower than in the *Ag class*. Activator contents are high for Mn (0·1 – 1·0%), lower for Cr and Co. Emission bands are narrow, or consist of broadened lines; some resolution into lines is often observed at low temperatures. Temperature dependence of efficiency is generally much less than in the *Ag class*, but current-saturation effects are more prominent for the conditions used.

Samarium occupies an intermediate position as an activator. Its constants are high for the *Mn class* (table 1, Part I), though its other properties fit best into this group.

The exponential processes require a more complex explanation than is provided by the usual mechanism assumed for crystal luminescence. It is suggested that the idea of electron traps might be applied to these phenomena. Energy dissipation has proved to be a useful parameter in correlation of results in different sections of the work, but for its adequate exploitation requires a knowledge of voltage and current changes in phosphors which we do not possess.

The simple hypothesis of linear voltage change in absorption accounts for some of the experimental results: in view of the complicated nature of luminescence, even so far as it is understood, such apparently simple explanations are not to be accepted too readily. There is a need for more experimental work on the lines of that recorded in these papers, and a more fundamental study of the electron-absorption process in phosphors.

ACKNOWLEDGMENTS

Our thanks are due to Mr. I. Shoenberg, of E.M.I., Ltd., for permission to publish this paper; and to Dr. W. Ehrenberg, formerly of the E.M.I. Research Laboratories, for helpful discussions, especially with regard to §2.

REFERENCES

ALLEN, 1932. *J. Franklin Inst.*, **214**, 25.
BROWN, 1937. *J. Opt. Soc. Amer.*, **27**, 186.
FANO, 1940. *Phys. Rev.*, **58**, 544.
HENDERSON, 1939. *Proc. Roy. Soc.*, A, **173**, 323.
KRÖGER, 1939. *Physica*, **6**, 369, 764, 779.
MOTT and GURNEY, 1940. *Electronic Processes in Ionic Crystals* (Oxford Univ. Press).
RANDALL, WILKINS, GARLICK, 1945. *Proc. Roy. Soc.*, A, **184**, 347–433.

THE INTEGRAL BREADTHS OF DEBYE-SCHERRER LINES PRODUCED BY DIVERGENT X RAYS

BY A. J. C. WILSON,
University College, Cardiff

MS. received 18 *December* 1945

ABSTRACT. Divergence of the incident X-ray beam produces appreciable phase differences between different parts of a crystal, even in the size range for which line broadening occurs. The broadening due to this phase difference is calculated. It is ordinarily negligible, as the increase in the width of the line is of the order of $1{\cdot}6\,t\cos\theta$, where t^3 is the volume of a crystal and θ is the Bragg angle. The special case of film and source equidistant from the crystal is investigated in greater detail.

§1. INTRODUCTION

In the theoretical treatment of the diffraction of X rays by crystals it is usual to consider a parallel incident beam. In practice, however, the rays diverge from a source a few centimetres from the specimen. The source may be real, the focus of the X-ray tube, or effective, part of the slit system of the camera. X-ray wave-lengths are sufficiently small for the divergence to introduce appreciable phase differences between different parts of a crystal small enough to produce line broadening. It seemed desirable, therefore, to investigate the diffraction of divergent x rays by crystals of this order of size, in case the divergence should have an appreciable effect on the line broadening.

It is found that the broadening on the film is of the order of $1{\cdot}6\,t\cos\theta$, where t^3 is the volume of a crystal and θ is the Bragg angle. To this approximation it is independent of the radius of the camera and the wave-length of the x rays. The broadening is thus negligible for crystals small enough to have their sizes measured by line broadening ($\sim 10^{-5}$ cm.), and it is not practically significant for the larger crystals used to give comparison lines. Debye-Scherrer lines would still have a finite breadth, due to the non-homogeneity of the incident x rays and imperfections in the experimental arrangement, even if diffraction broadening were entirely absent. Divergence broadening of the comparison lines would be indistinguishable from broadening due to the finite diameter of the specimen, and would disappear in the elimination of the broadening due to experimental conditions by methods such as that of Jones (1938).

The mathematical treatment is greatly simplified in the special case in which the source and the film are equidistant from the specimen. The calculation of the integral breadths of reflections from spherical crystals is carried through in detail for this case, so that the transition from small-particle (diffraction) broadening to large-particle (divergence) broadening can be followed. For any particular angle of reflection the integral breadth, β, is a minimum for a particle size given by $t = 1{\cdot}00(Q\lambda)^{\frac{1}{2}}/\cos\theta$, where Q is the camera radius and λ is the wave-length of the x rays. The broadening on the film, $Q\beta$, is $1{\cdot}40(Q\lambda)^{\frac{1}{2}}$. This is too small to be detected with normal technique.

§ 2. DIFFRACTION OF DIVERGENT X RAYS

In figure 1, the source of the x rays is at O, **P** is a vector joining the source to the centre of gravity of the crystal, **Q** is a vector from **P** to the point on the circumference of the camera at which it is desired to calculate the reflected intensity, $\mathbf{r}_j$ is a vector from **P** to the jth unit cell of the crystal, and 2θ is the angle of deviation.

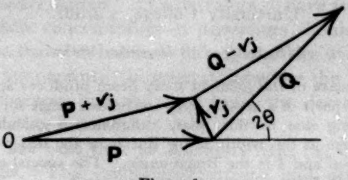

Figure 1.

The path difference between the rays diffracted from **P** and $\mathbf{P}+\mathbf{r}_j$ is

$$\begin{aligned}
&|\mathbf{P}+\mathbf{r}_j| - |\mathbf{P}| + |\mathbf{Q}-\mathbf{r}_j| - |\mathbf{Q}| \\
&\quad = \{(\mathbf{P}+\mathbf{r}_j)\cdot(\mathbf{P}+\mathbf{r}_j)\}^{\frac{1}{2}} - P + \{(\mathbf{Q}-\mathbf{r}_j)\cdot(\mathbf{Q}-\mathbf{r}_j)\}^{\frac{1}{2}} - Q \\
&\quad = P\left\{\left[1+\frac{2\mathbf{P}\cdot\mathbf{r}_j}{P^2}+\frac{\mathbf{r}_j\cdot\mathbf{r}_j}{P^2}\right]^{\frac{1}{2}}-1\right\} + Q\left\{\left[1-\frac{2\mathbf{Q}\cdot\mathbf{r}_j}{Q^2}+\frac{\mathbf{r}_j\cdot\mathbf{r}_j}{Q^2}\right]^{\frac{1}{2}}-1\right\}
\end{aligned}$$

$$=P\left\{\frac{\mathbf{P}\cdot\mathbf{r}_j}{P^2}+\frac{\mathbf{r}_j\cdot\mathbf{r}_j}{2P^2}-\frac{\mathbf{P}\cdot\mathbf{r}_j{}^2}{2P^4}+\dots\right\}+Q\left\{-\frac{\mathbf{Q}\cdot\mathbf{r}_j}{Q^2}+\frac{\mathbf{r}_j\cdot\mathbf{r}_j}{2Q^2}-\frac{\mathbf{Q}\cdot\mathbf{r}_j{}^2}{2Q^4}+\dots\right\}$$

$$=\mathbf{r}_j\cdot(\mathbf{p}-\mathbf{q})+\left(\frac{1}{2P}+\frac{1}{2Q}\right)\mathbf{r}_j\cdot\mathbf{r}_j-\mathbf{r}_j\cdot\left(\frac{\mathbf{pp}}{2P}+\frac{\mathbf{qq}}{2Q}\right)\cdot\mathbf{r}_j, \qquad \dots\dots(1)$$

where p and **q** are unit vectors in the direction of **P** and **Q**. Let $\mathbf{p}-\mathbf{q}=\lambda(\mathbf{h}+\boldsymbol{\eta})$, where **h** is the vector of the reciprocal lattice which makes $|\boldsymbol{\eta}|$ smallest. Then $\mathbf{r}_j\cdot\mathbf{h}$ is an integer and $\mathbf{r}_j\cdot\mathbf{h}\lambda$ can be dropped from the expression for the path difference. Let

$$\sigma=\frac{1}{2P\lambda}+\frac{1}{2Q\lambda},\quad \delta=\frac{1}{2P\lambda}-\frac{1}{2Q\lambda},\quad \rho=\sigma/\delta,\ \text{and}\ \boldsymbol{\Psi}=\frac{\mathbf{pp}}{2P\lambda}+\frac{\mathbf{qq}}{2Q\lambda}, \qquad \dots\dots(2)$$

where $\boldsymbol{\Psi}$ is a dyadic. The amplitude of the reflected beam is then (the "structure factor" being omitted for brevity)

$$G=\Sigma_j\exp\{2\pi i(\mathbf{r}_j\cdot\boldsymbol{\eta}+\sigma\mathbf{r}_j\cdot\mathbf{r}_j-\mathbf{r}_j\cdot\boldsymbol{\Psi}\cdot\mathbf{r}_j)\} \qquad \dots\dots(3)$$

$$\doteqdot V^{-1}\int\exp\{2\pi i(\mathbf{r}\cdot\boldsymbol{\eta}+\sigma\mathbf{r}\cdot\mathbf{r}-\mathbf{r}\cdot\boldsymbol{\Psi}\cdot\mathbf{r})\}\,dv, \qquad \dots\dots(4)$$

where V is the volume of one unit cell and the integration is over the volume of the crystal. This integral may be simplified somewhat by a proper choice of axes. Let

$$\mathbf{a}=(\mathbf{p}-\mathbf{q})/2\sin\theta,\quad \mathbf{b}=(\mathbf{p}+\mathbf{q})/2\cos\theta,\quad \boldsymbol{\gamma}=\mathbf{p}\times\mathbf{q}/\sin 2\theta. \qquad \dots\dots(5)$$

It will readily be seen that **a**, **b**, $\boldsymbol{\gamma}$ are unit and orthogonal. Then

$$\begin{aligned}\boldsymbol{\Psi}&=\sigma\sin^2\theta\mathbf{aa}+\delta\sin\theta\cos\theta\mathbf{ab}\\&\quad+\delta\sin\theta\cos\theta\mathbf{ba}+\sigma\cos^2\theta\mathbf{bb}.\end{aligned} \qquad \dots\dots(6)$$

It is desired to find unit and orthogonal axes $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ such that

$$\boldsymbol{\Psi}\cdot\boldsymbol{\alpha}=A\boldsymbol{\alpha},\quad \boldsymbol{\Psi}\cdot\boldsymbol{\beta}=B\boldsymbol{\beta}, \qquad \dots\dots(7)$$

where A and B are constants. It can be verified that

$$\left.\begin{aligned}\boldsymbol{\alpha}&=\frac{\mathbf{a}+\{|\rho\cot 2\theta|-|(\rho^2\cot^2 2\theta+1)^{\frac12}|\}\mathbf{b}}{\sqrt{2}\,|\{1-|\rho\cot 2\theta(\rho^2\cot^2 2\theta+1)^{\frac12}|+\rho^2\cot^2 2\theta\}^{\frac12}|},\\ \boldsymbol{\beta}&=\frac{\mathbf{a}+\{|\rho\cot 2\theta|+|(\rho^2\cot^2 2\theta+1)^{\frac12}|\}\mathbf{b}}{\sqrt{2}\,|\{1+|\rho\cot 2\theta(\rho^2\cot^2 2\theta+1)^{\frac12}|+\rho^2\cot^2 2\theta\}^{\frac12}|},\end{aligned}\right\} \qquad \dots\dots(8)$$

$$\left.\begin{aligned}A&=\tfrac12\{\sigma-|(\sigma^2\cos^2 2\theta+\delta^2\sin^2 2\theta)^{\frac12}|\},\\ B&=\tfrac12\{\sigma+|(\sigma^2\cos^2 2\theta+\delta^2\sin^2 2\theta)^{\frac12}|\},\end{aligned}\right\} \qquad \dots\dots(9)$$

and that $\boldsymbol{\alpha}\to\mathbf{a}$ and $\boldsymbol{\beta}\to\mathbf{b}$ as $\delta\to 0$. With axes in the directions $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$,

$$\mathbf{r}=x\boldsymbol{\alpha}+y\boldsymbol{\beta}+z\boldsymbol{\gamma},\quad \boldsymbol{\eta}=\xi\boldsymbol{\alpha}+\eta\boldsymbol{\beta}+\zeta\boldsymbol{\gamma},$$

$$G=V^{-1}\int\exp\{2\pi i[x\xi+y\eta+z\zeta+\sigma(x^2+y^2+z^2)-Ax^2-By^2]\}\,dv$$

$$=V^{-1}\int\exp\{2\pi i(x\xi+y\eta+z\zeta+Bx^2+Ay^2+\sigma z^2)\}dv \qquad \dots\dots(10)$$

$$\begin{aligned}&=V^{-1}\int\exp\{2\pi i[(B^{\frac12}x+\xi/2B^{\frac12})^2-\xi^2/4B]\}dx\int\exp\{2\pi i[(A^{\frac12}y+\eta/2A^{\frac12})^2\\&\qquad-\eta^2/4A]\}dy\int\exp\{2\pi i[(\sigma^{\frac12}z+\zeta/2\sigma^{\frac12})^2-\zeta^2/4\sigma]\}dz.\end{aligned} \qquad \dots\dots(11)$$

This integral is difficult to evaluate for an arbitrary crystal shape, but for a sphere whose radius a is large compared with $B^{-\frac{1}{2}}$, $A^{-\frac{1}{2}}$, $\sigma^{-\frac{1}{2}}$, it may be evaluated approximately by the use of properties of Fresnel integrals. It is well known that

$$\left.\begin{aligned} \left|\int_{x_1}^{x_2} \exp\{2\pi i c x^2\} dx\right|^2 &\sim (2c)^{-1} \quad \text{for } x_1 \text{ and } x_2 \text{ of opposite sign and } x_1^2,\ x_2^2 \gg c^{-1}, \\ &\sim 0 \qquad\quad \text{for } x_1 \text{ and } x_2 \text{ of the same sign.} \end{aligned}\right\} \quad \ldots\ldots(12)$$

For a sphere, the limits of integration of the three integrals in equation (11) are $\pm(a^2-y^2-z^2)^{\frac{1}{2}}$, $\pm(a^2-z^2)^{\frac{1}{2}}$, $\pm a$ respectively. The modulus of the value of the first integral in equation (11) is, therefore, approximately $(2B)^{-\frac{1}{2}}$ for $\xi^2<4B^2\{a^2-y^2-z^2\}$ (i.e. for $y^2<\{a^2-z^2-\xi^2/4B^2\}$), 0 otherwise. Similarly the modulus of the product of the first two integrals is $(4AB)^{-\frac{1}{2}}$ for $\eta^2<4A^2\{a^2-z^2-\xi^2/4B^2\}$ (i.e. for $z^2<\{a^2-\xi^2/4B^2-\eta^2/4A^2\}$) and zero otherwise, and the modulus of the triple integral is $(8AB\sigma)^{-\frac{1}{2}}$ for $\zeta^2<4\sigma^2\{a^2-\xi^2/4B^2-\eta^2/4A^2\}$ (i.e. for $\xi^2/4B^2+\eta^2/4A^2+\zeta^2/4\sigma^2<a^2$) and zero otherwise. Then the intensity of reflection as a function of $\boldsymbol{\eta}$ is

$$\left.\begin{aligned} H = VGG^* &= \frac{1}{8AB\sigma V}, \quad \frac{\xi^2}{4B^2}+\frac{\eta^2}{4A^2}+\frac{\zeta^2}{4\sigma^2}<a^2, \\ &= 0, \text{ otherwise.} \end{aligned}\right\} \quad \ldots\ldots(13)$$

In other words, the intensity of reflection is approximately constant and equal to $(8AB\sigma V)^{-1}$ within an ellipsoid of semi-axes of lengths $2Ba$, $2Aa$, $2\sigma a$ and directions $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, approximately constant and equal to zero outside. The total intensity I is equal to $(8AB\sigma V)^{-1}$ times the volume of the ellipsoid:

$$I = \tfrac{4}{3}\pi\,.\,2Ba\,.\,2Aa\,.\,2\sigma a\,.\,(8AB\sigma V)^{-1} = \tfrac{4}{3}\pi a^3 V^{-1},$$

which is the total number of unit cells, as it should be. The intensity with θ between θ and $\theta+d\theta$ is proportional to the volume of the ellipsoid contained between two planes perpendicular to $\mathbf{a}$ and a distance $2\cos\theta d\theta/\lambda$ apart. The maximum intensity is, therefore (figure 2),

$$\begin{aligned} \left(\frac{dI}{d\theta}\right)_0 &= \frac{2\cos\theta}{\lambda}\,.\,H\,.\,\pi\,.\,2\sigma a\,.\,T \\ &= \frac{\cos\theta}{2\lambda}\cdot\frac{\pi a T}{ABV}, \end{aligned} \quad \ldots\ldots(14)$$

where T is the semi-diameter of the ellipsoid in the direction of $\mathbf{b}$. The integral breadth is therefore

$$\beta = \frac{2I}{(dI/d\theta)_0} = \frac{16\lambda a^2 AB}{3T\cos\theta}. \quad \ldots\ldots(15)$$

The explicit expression for T is rather complex; it may be shown to be

$$T = a\left\{\frac{\boldsymbol{\beta}\cdot\mathbf{a}^2}{\dfrac{\boldsymbol{\alpha}\cdot\mathbf{a}^2}{4B^2}+\dfrac{\boldsymbol{\beta}\cdot\mathbf{a}^2}{4A^2}} + \frac{\boldsymbol{\alpha}\cdot\mathbf{a}^2}{\dfrac{\boldsymbol{\beta}\cdot\mathbf{a}^2}{4B^2}+\dfrac{\boldsymbol{\alpha}\cdot\mathbf{a}^2}{4A^2}}\right\}^{\frac{1}{2}}. \quad \ldots\ldots(16)$$

This simplifies considerably for $P=Q$. For most cameras this is approximately true, and the simplified expression will give some estimate of the integral breadth due to the divergence of the x-ray beam. The simplified expression is

$$T = 2Aa, \quad \ldots\ldots(17)$$

and the integral breadth becomes

$$\beta = \frac{8\lambda a B}{3\cos\theta} = \frac{8a}{3} \cdot \frac{\cos\theta}{Q} = 1{\cdot}65t\,\frac{\cos\theta}{Q}. \qquad \ldots\ldots(18)$$

The actual broadening on the film is equal to the camera radius times the integral breadth, i.e. to $1{\cdot}65\,t\cos\theta$. This becomes appreciable when a is a few hundredths of a millimetre.

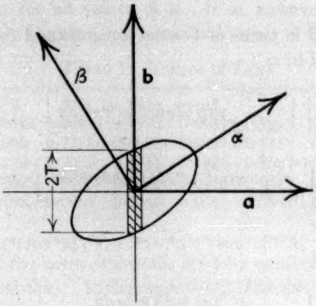

Figure 2.

The actual value of the numerical factor in equation (18) will depend on the shape of the crystal, and for non-spherical crystals on the indices of the reflection. Equation (11) can be evaluated similarly for other simple shapes; reflection from a face of a cubic crystal leads to a factor 2 instead of 1·65. Equation (18) should, however, give the order of magnitude of the effect.

§ 3. DETAILED CALCULATION FOR EQUIDISTANT SOURCE AND FILM

The whole calculation simplifies considerably for $P = Q$, so it is perhaps worth while to examine in more detail the transition from small-particle (diffraction) broadening to large-particle (divergence) broadening for a spherical particle in this special case. Equation (10) becomes

$$G = V^{-1}\int \exp\{2\pi i[x\xi + y\eta + z\zeta + \sigma(x^2\cos^2\theta + y^2\sin^2\theta + z^2)]\}dv, \qquad \ldots\ldots(19)$$

where both x and ξ are measured in the direction of $\mathbf{h}$. The intensity of reflection as a function of η is

$$H = VGG^* = V^{-1}\int_v\int_{v'} \exp\{2\pi i L\}\,dv\,dv' \text{ if}$$
$$L \equiv [\xi(x-x') + \eta(y-y') + \zeta(z-z') + (\{x^2 - x'^2\}\cos^2\theta + \{y^2 - y'^2\}\sin^2\theta + z^2 - z'^2)] \qquad \ldots\ldots(20)$$

so that

$$\frac{dI}{d\theta} = \frac{2\cos\theta}{\lambda}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H\,d\eta\,d\zeta$$
$$= \frac{2\cos\theta}{\lambda V}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_v\int_{v'} \exp\{2\pi i[L]\}dv\,dv'\,d\eta\,d\zeta. \qquad \ldots\ldots(21)$$

The integrals with respect to η and ζ are singular, being zero if $y' \neq y$ or $z' \neq z$, and infinite for $y' = y$ and $z' = z$, but "by an appropriate limiting process" (Patterson, 1939a, p. 973) it may be shown that the double integrals with respect to η and y', ζ and z' each have the value unity. Equation (21) becomes

$$\frac{dI}{d\theta} = \frac{2\cos\theta}{\lambda V}\iiiint \exp\{2\pi i[\xi(x-x') + \sigma(x^2 - x'^2)\cos^2\theta]\}\,dx\,dy\,dz\,dx', \qquad \ldots\ldots(22)$$

$$\left(\frac{dI}{d\theta}\right)_0 = \frac{2\cos\theta}{\lambda V}\iiiint \exp\{2\pi i\sigma\cos^2\theta(x^2 - x'^2)\}\,dx\,dy\,dz\,dx'. \qquad \ldots\ldots(23)$$

This may be expressed in terms of Fresnel integrals and cylindrical coordinates (axis in the direction of **h**) as

$$\left(\frac{dI}{d\theta}\right)_0 = \frac{2\cos\theta}{\lambda V}\int_0^{2\pi}\int_0^a\left[\int_{-(a^2-r^2)^{\frac{1}{2}}}^{(a^2-r^2)^{\frac{1}{2}}} \exp\{2\pi i\sigma\cos^2\theta x^2\}\,dx\right]$$

$$\left[\int_{-(a^2-r^2)^{\frac{1}{2}}}^{(a^2-r^2)^{\frac{1}{2}}} \exp\{-2\pi i\sigma\cos^2\theta x'^2\}\,dx'\right] r\,dr\,d\phi$$

$$= \frac{4\pi}{\lambda V\sigma\cos\theta}\int_0^a [C^2(2\sigma^{\frac{1}{2}}\cos\theta\sqrt{a^2-r^2}) + S^2(2\sigma^{\frac{1}{2}}\cos\theta\sqrt{a^2-r^2})]\,r\,dr$$

$$= \frac{\pi}{\lambda V\sigma^2\cos^3\theta}\int_0^{2\sigma^{\frac{1}{2}}a\cos\theta} [C^2(u) + S^2(u)]\,u\,du, \qquad \ldots\ldots(24)$$

where
$$C(u) + iS(u) = \int_0^u \exp\{\pi i u^2/2\}\,du.$$

The integral breadth is, therefore,

$$\beta = \frac{2I}{(dI/d\theta)_0} = 2\,.\,\frac{4\pi a^3}{3V}\,.\,\frac{\lambda V\sigma^2\cos^3\theta}{\pi}\,.\left[\int_0^{2\sigma^{\frac{1}{2}}a\cos\theta}[C^2(u) + S^2(u)]\,u\,du\right]^{-1}$$

$$= \frac{2\lambda}{3a\cos\theta}\,.\,\frac{1}{D(2\sigma^{\frac{1}{2}}a\cos\theta)}, \qquad \ldots\ldots(25)$$

where
$$D(u) = 4u^{-4}\int_0^u [C^2(u) + S^2(u)]\,u\,du. \qquad \ldots\ldots(26)$$

To progress further it is necessary to evaluate the function $D(u)$. Since $C^2(u) + S^2(u) \rightarrow u^2$ as $u \rightarrow 0$, its value for small values of u is 1. For large values of u, $C^2(u) + S^2(u) \rightarrow \frac{1}{2}$, so that the asymptotic value of $D(u)$ is u^{-2}. A series convenient for small values of u may be obtained as follows. It is known (Preston, 1895, p. 276) that

$$C^2(u) + S^2(u) = M^2 + N^2, \qquad \ldots\ldots(27)$$

where M and N are known series satisfying the differential equations

$$\frac{dM}{du} = 1 - \pi u N, \quad \frac{dN}{du} = \pi u M \qquad \ldots\ldots(28)$$

and, therefore,

$$\frac{d(M^2 + N^2)}{du} = 2M, \qquad \ldots\ldots(29)$$

and

$$D(u) = 2u^{-4}\int_0^{u^2}\left\{\int_0^{u^2} M u^{-1}\,d(u^2)\right\}d(u^2). \qquad \ldots\ldots(30)$$

Integrating the series for M (Preston, p. 275) term by term gives

$$D(u)=1-\frac{2\pi^2u^4}{3.4.1.3.5}+\frac{2\pi^4u^8}{5.6.1.3.5.7.9}-\frac{2\pi^6u^{12}}{7.8.1.3.5.7.9.11.13}+\dots \qquad \dots\dots(31)$$

The ratio of successive terms in this series is

$$-\frac{\pi^2u^4(2n-1)(2n)}{(2n+1)(2n+2)(4n-1)(4n+1)},$$

which approaches 0 for sufficiently large n for any value of u. It is therefore absolutely convergent for all values of u. It is, however, inconvenient for u greater than 2. Table 1 gives some numerical values of $D(u)$. Those for $u<2$ were calculated from the series, those for $u>2$ by numerical integration of four-place tables of $C(u)$ and $S(u)$. In the range 0·5–2·0 the greatest difference between the values calculated by the two methods is 0·0003; the mean difference is about 0·0001.

Table 1. Values of $D(u)$

	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0	1·0000	0·99999	0·99982	0·99911	0·99720	0·99317	0·98590	0·97405	0·95624	0·93101
1	0·8969	0·8534	0·8000	0·7373	0·6670	0·5917	0·5151	0·4413	0·3749	0·3190
2	0·2753	0·2440	0·2226	0·2077	0·1951	0·1821	0·1669	0·1508	0·1371	0·1257
3	0·1181	0·1118	0·1062	0·0992	0·0921	0·0856	0·0811	0·0774	0·0738	0·0695
4	0·0653	0·0628	0·0595	0·0569	0·0540	0·0512	0·0491	0·0472	0·0452	0·0430
5	0·0413									

With these values of $D(u)$, the integral breadth for any particular case can be calculated by equation (25). There is a minimum value of β which occurs for $D+uD'=0$, i.e. for $u \doteqdot 1{\cdot}24$. The corresponding value of a is

$$1{\cdot}24/2\sigma^{\frac{1}{2}}\cos\theta \sim 40{,}000\ \text{A. for } Q=10\ \text{cm.},\quad \lambda=2\ \text{A.},\quad \theta=45^\circ.$$

The minimum value of β is $1{\cdot}40\sqrt{\lambda/Q}$; the actual broadening on the film, $Q\beta$, is $1{\cdot}40\sqrt{Q\lambda}=0{\cdot}006$ mm., an amount undetectable with normal technique.

Equation (25) and the discussion in the previous paragraph have been given in terms of the radius of the particle, a. In terms of the cube root of its volume, t, the expressions become

$$\beta=\frac{1{\cdot}0747\,\lambda}{t\cos\theta}\cdot\frac{1}{D(1{\cdot}24\sigma t^{\frac{1}{2}}\cos\theta)}, \qquad \dots\dots(32)$$

$$t=1{\cdot}00/\sigma^{\frac{1}{2}}\cos\theta \text{ for minimum } \beta.$$

The minimum values of β and $Q\beta$ are unchanged. For $t/\sqrt{Q\lambda}$ small, equation (32) becomes $\beta=1{\cdot}0747\lambda/t\cos\theta$, in agreement with the results of Patterson (1939 b) and Stokes and Wilson (1942).

ACKNOWLEDGMENTS

The writer desires to thank Dr. H. Lipson, M.A., who suggested the desirability of investigating the effect of divergence on line breadths, for his encouragement and interest throughout the work, and Dr. P. P. Ewald for helpful criticisms.

REFERENCES

Jones, F. W., 1938. *Proc. Roy. Soc.*, A, **166**, 16.

Patterson, A. L., 1939 a. *Phys. Rev.*, **56**, 972; 1939 b. *Ibid.*, **56**, 978.

Preston, T., 1895. "The Theory of Light" (second edition) (London: Macmillan and Co.).

Stokes, A. R. and Wilson, A. J. C., 1942. *Proc. Camb. Phil. Soc.*, **38**, 313.

THE LIGHT-DIFFUSING PROPERTIES OF MAGNESIUM OXIDE

By V. G. W. HARRISON,
The Printing and Allied Trades Research Association, London

MS. received 28 *December* 1945; *in revised form* 29 *March* 1946

ABSTRACT. Measurements made on magnesium oxide deposited to a depth of 2 mm. on aluminium show that, for angles of incidence up to 60°, the polar reflexion curves are in general *ellipses*, and not the circles required by Lambert's law. For angles of incidence less than 45°, the major axis of the ellipse is perpendicular to the plane of the test surface, while for angles of incidence from 45° to 60° the minor axis of the ellipse is perpendicular to this plane. At 45° incidence, the ellipse degenerates to a circle, and in this special case only, Lambert's law is obeyed to within the accuracy of the measurements. For angles of incidence greater than 60°, the curves become quite irregular, and at almost grazing incidence there is a distinct reddish specular component in the reflected light. An attempt is made to analyse the curves into specular and diffuse components by Barkas's method, but the results suggest that this method is not really applicable in the present instance. 45° appears to be a suitable angle of incidence to adopt in practical gloss determinations.

§1. INTRODUCTION

THOSE interested in measuring the gloss of real surfaces frequently need for reference a reproducible matt surface. The ideal matt surface is unknown; nevertheless there are several conveniently-prepared real surfaces which approximate to it fairly closely under certain conditions of illuminating and viewing; and of these, the surface of magnesium oxide deposited on silver or aluminium, widely used as a standard in colorimetry, merits special consideration. Not only is such a surface easily prepared and of high reflecting power (97–98%), but it is also as nearly a true matt as is any other surface that has so far been suggested.

If, in figure 1, XY represents the ideal matt surface illuminated by a beam IO incident at an angle i, then the intensity of the light scattered in the direction OP is, by definition, proportional to $c \,.\, \cos i \,.\, \cos R$, where c is a factor defining the brightness of the surface and R is the angle of viewing. More exactly, the *illumination* at the point P per unit area of the diffusing surface is given by the equation

$$L=(cI/d^2)\,.\,\cos i\,.\,\cos R,$$

where I is the intensity of the incident beam and d is the distance of the point P from O. This relation is the well-known Lambert's law (Lambert, 1760). It is assumed that the dimensions of the diffusing surface are small compared with the distance d. It follows that if I_0 is the intensity of the radiation emitted normally to the surface, then the intensity of the radiation emitted at any other angle R is, for constant conditions of illumination, given by the simple equation

$$I_R=I_0\,.\,\cos R.$$

The apparent simplicity of Lambert's law is deceptive; although quite a number of surfaces obey it approximately under certain conditions of illuminating and viewing, no material is known for which it holds exactly, and none of the several attempts to account for it theoretically has been very successful. The truth of the law is certainly not self-evident, and its value lies only in its convenience and simplicity, since it is the only relation for which the brightness of the radiating or reflecting surface is independent of the angle of viewing. Bouguer (1762) tried to account for Lambert's law on the supposition that the reflecting surface behaves as an aggregate of small mirrors set at all possible inclinations. Lommel (1880) considers the case of glowing masses of metals and gases which emit light not only from their surface, but also throughout their whole volume, and he later attempts to extend this theory to account for the reflecting properties of matt surfaces (1889). Unfortunately the equations derived are of little practical value.

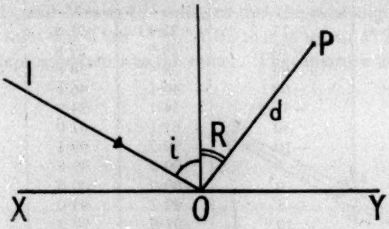

Figure 1.

Further work on this subject has been done by Grabowski (1914), Chinmayanandam (1919), Berry (1923), Pokrowski (1924 to 1926), Schulz (1925) and Barkas (1939). Attempts to discover materials for which Lambert's law is obeyed have been made by Bouguer (1762), Provostaye and Desains (1851), Kononowitsch (1879 and 1881) and Möller (1885). Ångström (1885) found that the polar light-intensity distribution curves were ellipses rather than the circles required by Lambert's law: for the substances examined, he found that the curves were sections of prolate spheroids for small angles of incidence, and oblate spheroids for large angles of incidence; and at an angle of incidence of about 30° the spheroids degenerated into a sphere, so that at this particular angle of incidence the simpler form of Lambert's law is obeyed. Further studies of matt surfaces have been made by Godard (1887), Seeliger (1888), Messerschmidt (1888), Wiener (1892), Jenko (1898), Wright (1900), Thaler (1903), Henning and Heuse (1922) and Woronkoff and Pokrowski (1924),.

§2. EXPERIMENTAL DETAILS

(a) *Preparation of specimens*

In order to get a reproducible standard surface for colour and gloss measurements, the magnesium-oxide layer should be infinitely thick. In practice, this is taken to be a layer which is so thick that doubling its thickness produces no appreciable change in its reflecting properties. Unfortunately, magnesium oxide is not as opaque as could be desired, and a deposit of several millimetres must be made before the layer can be regarded as "infinitely" thick. As the thickness of the layer increases, the tendency for the deposit to fall off its support in large flakes also increases rapidly, and in practice it becomes difficult to work with a layer more than 2 or 3 mm. in thickness; some compromise must therefore be made. For the experiments here described, the diffusing surfaces were prepared by holding a disc of aluminium, 73 mm. in diameter and 5 mm. in thickness, at a distance of about 10 cm. above a piece of burning magnesium ribbon. The process of burning the ribbon was continued until the thickness of the deposit reached 2 mm. In order to check whether this thickness was sufficient to reduce the effect of the aluminium support to negligible proportions, goniophotometric measurements were made on layers of 1 mm. and 2 mm. thickness, with the results shown in table 1. It will be seen that, while doubling the thickness of the layer has increased the diffusion slightly (as shown by the slightly greater intensity of light scattered at large angles to the normal), the increase is nevertheless quite small, and a further increase in the thickness of the deposit would not cause appreciable differences in the form of the polar curves.

Table 1. Effect of the thickness of the magnesium oxide layer on the intensity of the light scattered in various directions (for angle of incidence -30°)

Angle of viewing (degrees)	Relative intensity for layer of thickness	
	1 mm.	2 mm.
−75	21·4	21·8
−70	29·6	29·8
−65	38·4	38·6
−60	46·1	46·6
−55	54·1	54·9
−50	61·1	62·0
−10	99·2	99·1
− 5	100·0	99·8
0	100·0	100·0
5	99·2	99·0
10	97·8	97·3
15	95·3	95·1
20	92·0	91·9
25	87·9	88·0
30	83·2	83·3
35	77·6	78·9
40	71·3	72·3
45	65·8	66·7
50	59·0	59·7
55	51·8	53·1
60	43·7	44·7
65	36·1	37·2
70	28·0	29·2
75	20·5	21·2

(b) *Technique of measurement*

All measurements were made on the new Patra Hilger Goniophotometer, an instrument following the general lines of the model described in Patra Research

Report No. 1 (July 1938—now out of print). Particulars of this instrument are to be published shortly. The essential features are given in figure 2. Light from a ribbon filament lamp is collected by a condenser L_1, and the optical system is such that an image of the slit S_1 is thrown on to the test specimen M, the length of the slit being parallel to the axis about which the specimen and photometer arm rotate. The angular spread of the incident beam is controlled mainly by the slit S_2: the smaller the opening, the more nearly parallel is the incident light, but the intensity of the light reaching the photocell P is correspondingly reduced. In practical measurements one has always to effect a compromise between a reasonably small angular aperture and workable galvanometer deflexions. The photocell P is of the ordinary rectifier type connected to a sensitive galvanometer: the angular spread of the wedge of light reaching this receptor is controlled by the slit S_3.

In the arrangement used, the width of the band of light leaving the back surface of the lens L_3 was 12 mm., and the width of the slit image at M was 1 mm. The distance between L_3 and M was 185 mm., so that the total angular spread of the wedge of incident light was $2 \tan^{-1}$ (55/1850), that is, about 3° 24′. The angular aperture of the receptor system was the same. This aperture would be too great for materials possessing " peaky " goniophotometric curves and would have to be cut down; but with surfaces of very low gloss it does not introduce serious errors into the form of the curves. More important from the present point of view is the size of the spot of light thrown on to the specimen. It is assumed in all these measurements that the dimensions of the spot of light on M are small compared with the distance of the photocell from the test surface (PM). If this is not so, the curves can be appreciably distorted. In the arrangement used, measurements were made only in the plane containing the angle of incidence, and the length of the spot of light (being at right angles to this plane) was thus of minor importance and was kept to 14 mm. The width of the slit S_1 was reduced until the width of the rectangular spot of light thrown on to M was 1 mm. for normal incidence (for oblique angles of incidence the width of the slit image becomes correspondingly greater). The distance from L_4 to M in the instrument used was 17 cm. and from L_4 to P, 20 cm. The focal length of the lens L_4 was about $18\frac{1}{2}$ cm.

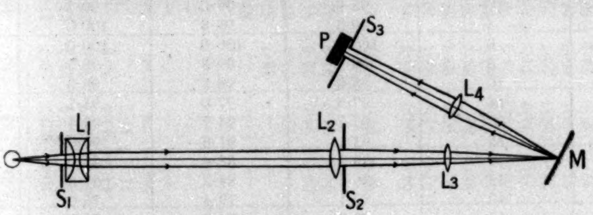

Figure 2.

To show the effect of the size of the spot of light on the form of the curves, the figures in table 2 are reproduced. Those in columns 2 and 3 were taken with the goniophotometer, now destroyed, described in Patra Research Report No. 1. This instrument used a photocell of about $3\frac{3}{4}$ cm. diameter placed at a distance of 50 cm. from the test surface; no lens was used in the receptor system. A circular spot of light was thrown on to the specimen (that is, at normal incidence, the circle being broadened into an ellipse for oblique incidence), and this circle was usually of 1 cm. diameter: the figures given by this arrangement are in column 2. The corresponding figures given when the spot of light was cut down to 4 mm. diameter are in column 3. The effect on the curves is clearly seen. Finally, in column 4 are the figures obtained with the present instrument, with the rectangular spot of light 14 × 1 mm. The figures are all corrected so as to give a reading of 100·0 when the angle of viewing is 0°. The broadening of the curves with large spots of light is attributed to the fact that with a large spot of light the effective area sending flux to the photocell tends to increase as the angle of viewing increases.

Table 2. Effect of the dimensions of the spot of light thrown on to the test specimen (for angle of incidence −45°)

Angle of viewing (degrees)	Apparent relative intensity for spot of light		
	10 mm. diam. circle	4 mm. diam. circle	14 × 1 mm. rectangle
−75	25·5	22·7	22·9
−70	34·8	31·4	31·4
−65	43·8	40·0	40·2
−25	93·0	91·3	90·9
−20	95·4	94·7	94·9
−15	97·5	97·1	97·1
−10	98·8	99·0	99·3
− 5	99·5	99·8	100·0
0	100·0	100·0	100·0
5	99·5	99·8	99·3
10	98·8	98·8	98·0
15	97·5	97·0	95·1
20	95·7	94·7	93·4
25	93·1	91·8	90·0
30	89·9	88·4	86·0
35	86·1	84·2	81·8
40	82·1	79·4	76·0
45	77·2	74·0	70·1
50	71·9	68·2	63·8
55	65·3	61·8	57·1
60	58·3	54·8	49·7
65	50·8	47·3	41·9
70	42·1	39·2	34·1
75	32·7	30·6	25·0

§ 3. RESULTS

The results of the measurements on magnesium oxide are given in table 3, which shows the measured intensity for various angles of incidence and viewing, expressed as a percentage of the maximum intensity obtainable with the given settings of the instrument. This maximum intensity occurs when the directions of incidence and viewing are both normal to the test surface. The intensity cannot be measured experimentally under these conditions, but it can be determined to a satisfactory degree of accuracy by extrapolation, since it is found that

Table 3. Relative intensity of the light scattered from a surface of magnesium oxide at various angles of incidence and viewing

Angle of viewing	Angle of incidence															
	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
−75	19·8	20·1	20·4	20·1	20·4	20·4	20·4	20·2	20·7	21·1	21·7	23·1	—	—	—	—
−70	27·9	27·2	27·6	27·6	27·2	27·6	27·9	27·9	28·2	28·9	28·5	—	—	—	—	—
−65	35·0	34·6	35·6	35·3	35·0	35·3	35·6	35·7	36·0	37·0	—	—	—	—	—	—
−60	42·1	43·1	43·1	43·1	42·7	44·1	43·5	43·4	43·8	—	—	—	—	—	—	—
−55	49·8	50·9	51·2	50·6	50·6	52·2	51·6	51·2	—	—	—	—	—	—	—	48·1
−50	57·1	58·0	58·8	58·3	57·8	59·0	59·1	—	—	—	—	—	—	—	53·2	52·6
−45	63·6	65·5	65·5	64·8	65·5	66·1	—	—	—	—	—	—	—	61·0	57·7	56·8
−40	70·8	71·8	72·7	71·9	72·0	—	—	—	—	—	—	—	66·8	64·8	61·9	60·8
−35	76·1	77·3	79·1	77·8	—	—	—	—	—	—	—	73·2	70·8	69·4	65·6	64·2
−30	81·7	83·0	84·2	—	—	—	—	—	—	—	78·7	77·1	74·7	72·8	68·8	66·8
−25	86·5	87·7	—	—	—	—	—	—	—	83·8	82·3	81·1	77·3	75·8	71·8	69·4
−20	91·4	—	—	—	—	—	—	—	88·7	87·4	85·0	84·8	80·1	77·8	74·7	71·4
−15	—	—	—	—	—	—	—	92·1	91·4	89·7	87·7	86·7	82·3	80·1	76·0	73·2
−10	—	—	—	—	—	—	94·1	94·0	92·8	91·6	89·7	88·2	83·8	81·8	77·6	74·7
−5	—	—	—	—	—	96·5	94·7	94·7	93·9	92·2	90·9	88·7	84·8	82·2	78·6	75·8
0	(100)	—	—	—	96·7	96·7	95·0	94·7	94·1	92·2	91·2	88·7	85·1	82·8	79·2	75·9
5	—	—	—	96·7	96·0	96·0	94·2	94·7	93·5	91·6	90·2	88·2	85·1	82·4	79·2	75·9
10	—	—	97·2	95·2	94·7	94·7	93·0	93·0	92·1	90·1	88·8	87·0	84·0	82·2	78·9	75·9
15	—	94·0	95·2	92·7	92·1	92·1	90·9	90·8	89·7	87·7	86·7	85·7	82·3	81·1	78·2	76·1
20	92·1	90·3	92·1	88·7	89·5	89·7	88·2	86·7	86·2	86·1	84·8	83·7	81·8	79·8	77·2	75·2
25	87·2	86·2	86·9	84·2	85·7	85·7	84·3	83·1	82·9	83·1	81·8	81·1	79·6	77·9	75·8	74·9
30	82·2	81·1	82·2	79·8	80·8	81·1	80·4	79·2	79·2	79·1	78·5	78·2	77·1	75·9	74·7	73·3
35	76·6	75·8	76·5	74·6	75·2	76·1	75·2	74·7	74·7	75·2	74·7	74·4	73·9	73·3	72·2	71·8
40	70·4	69·7	70·0	68·8	70·0	70·8	70·1	69·3	69·7	70·1	70·1	70·1	70·1	70·8	70·4	70·8
45	63·6	63·2	63·9	62·6	63·6	64·2	63·9	64·1	63·9	64·6	65·8	66·2	66·2	69·7	68·2	68·8
50	56·3	55·8	57·0	55·1	56·3	57·3	57·3	58·0	58·1	58·7	60·3	61·7	62·3	64·6	65·7	67·1
55	48·6	48·6	49·5	48·0	49·3	49·9	50·7	51·2	51·7	52·6	54·4	56·1	57·8	61·0	62·7	66·2
60	40·8	40·9	41·5	41·5	41·5	42·5	43·1	44·4	44·4	45·7	48·0	50·3	52·7	56·4	59·8	64·2
65	33·1	32·7	33·7	33·7	33·7	35·0	35·6	36·4	37·3	38·6	41·2	43·5	46·4	51·7	57·1	62·2
70	24·9	25·3	25·9	25·9	26·3	27·3	27·9	28·9	29·2	31·4	33·4	36·0	39·2	45·1	52·9	62·8
75	17·5	17·8	18·2	17·8	18·2	19·1	20·1	20·7	21·4	23·0	25·0	27·9	30·8	36·4	44·1	60·3

the experimental curves are very nearly ellipses. The figures are in fairly good agreement with those of Barkas (1939).

Examination of these results shows at once that the variation with i predicted by Lambert's law is not exactly true. If it were, the *flux* reaching the photocell would be independent of the angle of incidence. We find, in fact, that the flux received by the photocell is only approximately constant and independent of the angle of incidence when the angle of viewing is at the special angle of 40°. The simpler form of Lambert's law ($I_R = I_0 \,.\, \cos R$) holds reasonably well for an angle of incidence of 45°, but it does not hold accurately at any other angle of incidence.

The polar curves for angles of incidence less than 45° are found to be ellipses with the major axis perpendicular to the test surface. The polar equation of an ellipse with origin of polar co-ordinates at one end of the major axis is

$$I = \frac{2ab^2 \,.\, \cos R}{a^2 \sin^2 R + b^2 \cos^2 R},$$

where a and b are the semi-major and semi-minor axes. By a suitable choice of a and b, this equation expresses the experimental results to within the accuracy of measurement. Table 4 shows the measured and calculated values for angles of incidence of 0°, 15°, 25°, 30°, 40°, 45°, 50°, 55° and 60°. It will be seen that in most cases the agreement is good. For angles of incidence less than 45° the curves thus represent sections of a prolate spheroid, as recorded by Ångström (1885). At 45° incidence the ellipse degenerates to a circle. For angles of incidence between 45° and 55° the curves are approximately sections of an oblate spheroid—that is, they are again ellipses, but this time with the *minor* axis perpendicular to the plane of the surface. At 60° incidence, the ellipse begins to lose its symmetry and develops a " bulge " in the region of 60° to 70°; for angles of incidence of 65° and over, the curves become quite irregular and no simple equation can be found to fit them; and at nearly grazing incidence (80° and over) there is distinct regular reflexion from the surface, which, at the appropriate angle of viewing, looks polished with a reddish tinge. The polar curves for 0°, 45° and 55° incidence are drawn in figure 3.

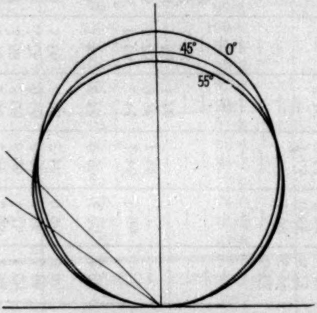

Figure 3.

Table 4. **Comparison of observed values of intensity with those calculated from the ellipse** $I=\dfrac{2ab^2\cos R}{a^2\sin^2 R+b^2\cos^2 R}$

	Angle of viewing									
	0	10	20	30	40	50	60	70	80	90
$i=0^{\circ}$, $a=50{\cdot}0$, $b=45{\cdot}0$										
I, calculated	100	97·8	92·4	82·0	69·7	56·5	42·7	28·4	14·2	0
I, observed			92·1	82·2	70·4	56·3	40·8	24·9		
			91·4	81·7	70·8	57·1	42·1	27·9		
$i=15^{\circ}$, $a=48{\cdot}7$, $b=44{\cdot}5$										
I, calculated	97·4	95·5	90·0	80·2	68·9	56·1	42·4	28·4	14·2	0
I, observed		95·2	88·7	79·8	68·8	55·1	41·5	25·9		
					71·9	58·3	43·1	27·6		
$i=25^{\circ}$, $a=48{\cdot}3$, $b=45{\cdot}4$										
I, calculated	96·7	94·9	90·0	80·9	69·9	57·2	43·8	29·5	9·9	0
I, observed	96·7	94·7	89·7	81·1	70·8	57·3	42·5	27·3		
						59·0	44·1	27·6		
$i=30^{\circ}$, $a=47{\cdot}5$, $b=44{\cdot}6$										
I, calculated	95·0	93·1	88·1	79·3	69·0	57·0	43·4	29·3	14·7	0
I, observed	95·0	93·0	88·2	80·4	70·1	57·3	43·1	27·9		
		94·1				59·1	43·5	27·9		
$i=40^{\circ}$, $a=47{\cdot}0$, $b=45{\cdot}5$										
I, calculated	94·2	92·3	88·0	80·0	69·8	58·2	44·5	30·3	15·5	0
I, observed	94·1	92·1	86·2	79·2	69·7	58·1	44·0	29·2		
		92·8	88·7				43·8	28·2		
$i=45^{\circ}$, $a=46{\cdot}1$, $b=46{\cdot}1$										
I, calculated	92·2	90·8	86·6	79·9	70·6	59·2	46·1	31·6	16·0	0
I, observed	92·2	90·1	86·1	79·1	70·1	58·7	45·7	31·4		
		91·6	87·4					28·9		
$i=50^{\circ}$, $a=45{\cdot}6$, $b=46{\cdot}0$										
I, calculated	91·2	89·8	86·2	79·2	70·4	59·4	46·1	31·8	16·1	0
I, observed	91·2	88·8	84·8	78·5	70·1	60·3	48·0	33·4		
		89·7	85·0	78·7				28·5		
$i=55^{\circ}$, $a=44{\cdot}4$, $b=46{\cdot}8$										
I, calculated	88·7	87·9	84·6	78·9	71·0	60·9	48·0	33·4	17·2	0
I, observed	88·7	87·0	83·7	78·2	70·1	61·7	50·3	36·0		
		88·2	84·8	77·1						
$i=60^{\circ}$, $a=42{\cdot}6$, $b=47{\cdot}5$										
I, calculated	85·4	84·4	82·0	77·2	70·8	62·0	49·8	35·4	18·4	0
I, observed	85·1	84·0	81·8	77·1	70·1	62·3	52·7	39·2		
		83·8	80·1	74·7	66·8					

No theoretical explanation is here advanced as to why the polar curves should be ellipses rather than circles. These results are presented merely as experimental facts. Some other expressions were examined to see whether the experimental results would fit them better than the equation of the ellipse. Thus the equation

$$I = I_0 \cos R \, . \, e^{-Kt.\sec R}$$

holds approximately for the light transmitted through certain diffusing layers (Harrison, 1940), where t is the thickness of the layer and K is a constant. The equation gives families of oval curves. No values of Kt could be found which gave better agreement between observed and calculated values than is shown in table 4. The equations of Henning and Heuse (1922), Lommel (1880, 1889) and Ångström (1885) were also examined, but no real values of the constants could be found such that these equations would fit the observed results.

§ 4. THE BARKAS ANALYSIS

The results were examined by the Barkas analysis (1939), in which the reflecting properties of the surface are expressed in terms of an optically equivalent surface made up of mirror facets and rough facets, tilted at all angles. The fundamental equation may be written (in our present notation) as

$$I_{i,R} \, . \cos i = K\{\cos(i-R) + \cos(i+R)\} + sB \, . \cos \tfrac{1}{2}(i-R),$$

where K is a factor defining the contribution of the rough elements, while B is a similar factor defining the contribution of the mirror elements. s is a function of $\frac{1}{2}(i-R)$ and is calculated from Fresnel's laws of reflexion. Both K and B are in general functions of i and R.

Table 5. Values of K [Barkas's $\frac{1}{2}(rA)$] for increasing values of $(i+R)$

$(i+R)$	K
0	39·8
5	40·4
10	40·1
15	41·3
20	42·9
25	42·4
30	42·4
35	42·8
40	41·5
45	41·8
50	40·7
55	41·2
60	38·8
65	36·1
70	(37·6)
75	(33·4)

In accordance with Barkas's treatment, $I_{i,R} \, . \cos i$ was plotted against $(i-R)$, the curves obtained being closely similar to, but not identical with, those published by Barkas. The values of K [Barkas's $\frac{1}{2}(rA)$] were calculated from the curves by his method over the range $(i+R) = 0°$ to $75°$, with the results shown in table 5. The last two values (in parentheses) cannot be regarded as very accurate, and it is not possible to determine K for values of $(i+R)$ greater than 75° by Barkas's method:

there is, however, clear evidence that the value of K is beginning to fall fairly rapidly for values of $(i+R)$ greater than 50°. Barkas quotes a value of $0{\cdot}41_5$ for K for all values of $(i+R)$ up to 70°, and he assumes that this value of K holds good when $(i+R)$ is greater than 70° (when K cannot be determined directly). Examination of table 5 will show that, while the mean value of K over the whole range (40·4) is in satisfactory agreement with Barkas's value, nevertheless the figures do show a decided trend, and suggest that K rises to a maximum at about $(i+R)=35°$, and thereafter diminishes with increasing rapidity.

Table 6. Values of B for $K=0{\cdot}41$ (assumed constant)

$(i-R)$	$(i+R)$				
	0	10	20	30	40
0	2·5	2·2	1·9	1·5	0·9
10	2·2	2·0	1·7	1·3	0·7
20	1·7	1·6	1·2	0·8	0·3
30	1·2	1·2	0·6	0·0	−0·1
40	0·9	0·9	0·3	−0·2	−0·2
50	0·8	0·6	0·1	−0·3	—
60	0·9	0·8	0·2	—	—
70	1·0	—	—	—	—

There are further difficulties. If, despite the evidence, one assumes that K is, in fact, constant over the whole range, and one calculates the values of B according to the method which Barkas has outlined in his paper, one gets the results in table 6. These values are in satisfactory agreement with those given by Barkas, but there are some negative values. This does not seem plausible; moreover, the table suggests that in the cases where it has not been possible to determine B experimentally, it may have even greater negative values than are here recorded. On the other hand, if we now, instead of assuming that K is constant, take the experimental values given in table 5, we get the values of B given in table 7. Here the values of B do not agree with those of Barkas, and there is a marked tendency for them to rise again after passing through a minimum near $(i+R)=20°$. When plotted, they give the curve shown in figure 4, which expresses the comparative areas of the mirror facets at various

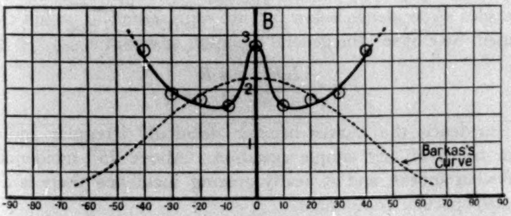

Figure 4.

inclinations to the mean surface: the form of this curve is a good deal less plausible than the one given by Barkas (shown dotted), though one can hardly say definitely that it is wrong. The truth seems to be that the Barkas analysis

Table 7. Values of B for K varying between 42·9 and 33·4, according to table 5

$(i-R)$	$(i+R)$				
	0	10	20	30	40
0	2·8	1·7	1·8	1·9	2·7
10	2·5	1·5	1·6	1·7	2·5
20	2·0	1·1	1·1	1·2	1·8
30	1·5	0·8	0·5	0·6	1·0
40	1·4	0·2	0·2	0·0	0·2
50	1·0	0·3	0·0	0·0	0·0
60	1·0	0·5	0·2	—	—
70	1·2	—	—	—	—

is not very satisfactory when applied to surfaces as matt as magnesium oxide, and unless K can be determined accurately for a much wider range of values of $(i+R)$ the values of B will remain open to considerable doubt.

§ 5. SUMMARY AND CONCLUSIONS

It has been shown that Lambert's law does not hold for a surface of magnesium oxide 2 mm. thick, since the flux emitted per unit area is not proportional to the cosine of the angle of incidence. The simpler form of the law,

$$I_R = I_0 \,.\, \cos R,$$

is valid only in the special case where the angle of incidence is 45°. The polar curves for angles of incidence below 60° are represented to within the accuracy of measurement by ellipses, the general equation of which is

$$I = \frac{2ab^2 \,.\, \cos R}{a^2 \sin^2 R + b^2 \cos^2 R}.$$

Above 60° incidence the curves become decidedly irregular and cannot be expressed in terms of any simple equation. Above 75° incidence signs of specular reflexion appear, and at nearly grazing incidence there is a fairly well defined reddish specular component.

This is not the place to discuss the design of glossmeters. In practice, one finds instruments making use of angles of incidence from 45° to over 80°. Nevertheless, the fact that 45° incidence is the only angle for which Lambert's law in its simpler form holds for magnesium oxide seems to be a fairly powerful argument for the adoption of 45° incidence in practical work, particularly as 45° is a very convenient angle from the point of view of instrument design. The point is, with magnesium oxide at 45° incidence we have a conveniently prepared surface which does give a polar curve approximating closely to Lambert's circle, while at 60° incidence and more, it is doubtful whether there is any real surface for which Lambert's law is even approximately true. At 45° incidence, therefore, we have a useful reference material of zero gloss, while at 60° incidence and over, we have none.

ACKNOWLEDGMENT

Finally, my thanks are due to the Council of the Printing and Allied Trades Research Association for permission to publish this paper.

REFERENCES

Ångström, K., 1885. *Ann. Phys., Lpz.*, **26**, 253.

Barkas, W. W., 1939. *Proc. Phys. Soc.*, **51**, 274.

Berry, E. M., 1923. *J. Opt. Soc. Amer.*, **7**, 627.

Bouguer, P., 1762. *Traité d'optique* (Richtenburg's Latin translation, Book 2, Sect. 3, Art. 3).

Chinmayanandam, T. K., 1919. *Phys. Rev.*, **13**, 96.

Godard, L., 1887. *Ann. Chim. (Phys.)*, **10**, 354.

Grabowski, L., 1914. *Astrophys. J.*, **39**, 299.

Harrison, V. G. W., 1940. *Proc. Tech. Sect. Paper Makers' Assoc.*, **21**, 67.

Henning, F. and Heuse, W., 1922. *Z. Phys.*, **10**, 111.

Jenko, P., 1898. *Ann. Phys., Lpz.*, **66**, 1182.

Kononowitsch, A., 1879. *Fortschr. Phys.*, **35**, 430 ; 1881. *Ibid.*, **37** (2), 481.

Lambert, J. H., 1760. *Photometria* (Augsburg), Part 1, Cap. 2, Sections 53, 72, 77–81.

Lommel, E., 1880. *Ann. Phys., Lpz.*, **10**, 449 ; 1889. *Ibid.*, **36**, 473.

Messerschmidt, J. B., 1888. *Ann. Phys. Lpz.*, **34**, 867.

Möller, W., 1885. *Ann. Phys., Lpz.*, **24**, 266.

Pokrowski, G. I., 1924. *Z. Phys.*, **30**, 66 ; 1925. *Ibid.*, **35**, 34 ; 1926. *Ibid.*, **35**, 390 and 472.

de la Provostaye, F. and Desains, P., 1851. *C.R. Acad. Sci., Paris*, **33**, 444. Also 1847. *Ibid.*, **24**, 60 ; *Ibid.*, 1848. **26**, 212 ; 1852. *Ann. Chim. (Phys.)*, **34**, 192.

Schulz, H., 1925. *Z. Phys.*, **31**, 496.

Seeliger, H., 1888. *Sitzungsberichte der math. phys. Klasse der Königl. bayerischen Akad. der Wissenschaften zu München*, **18**, 201.

Thaler, F., 1903. *Ann. Phys., Lpz.* (4), **11**, 996.

Wiener, C., 1892. *Ann. Phys., Lpz.*, **47**, 638.

Woronkoff, G. P. and Pokrowski, G. I., 1924. *Z. Phys.*, **20**, 358.

Wright, H. R., 1900. *Phil. Mag.*, **49**, 199.

THE BEHAVIOUR OF WATER UNDER HYDROSTATIC TENSION: I

By H. N. V. TEMPERLEY and LL. G. CHAMBERS,
Admiralty Research Laboratories

MS. received 3 *December* 1945

ABSTRACT. A study is made of the behaviour of water under hydrostatic tension. A review of the literature reveals very considerable differences in the values of critical tension obtained by different methods, the factor of disagreement being sometimes as high as 50. Two methods are studied in detail: the Berthelot method, which makes use of the difference in expansions of glass and water and the Reynolds centrifugal method. Evidence is brought forward that it requires high pressures to persuade water to stick to glass under subsequent tension. A modification of the Berthelot method gives results in much better agreement with other methods. Examination of the Reynolds method shows that account must be taken of the fact that water does not move as a rigid body. Although this has no appreciable effect on the pressure distribution, it does imply a considerable stirring of the water, and any small bubbles would tend to be brought near the region of greatest tension in a time that might be comparable with that of the experiment.

The flow of water in constricted tubes is also studied. The critical tension appears to be low, but this is to be expected in view of the fact that the motion is probably turbulent before cavitation occurs.

It is concluded that if tension is applied statically, ordinary water can stand tensions of the order of 40 atmospheres, even if it is not perfectly air-free. Water nearly saturated with air has been shown to stand tension up to 6 atmospheres.

---

§ 1. INTRODUCTION

It has long been known that, under suitable conditions, liquids are capable of sustaining considerable tensions. While under tension, the liquid is in a metastable state and, if the tension is too great, it changes irreversibly to the two-phase system liquid and vapor and dissolved gas. The critical tension is of interest in quite a number of fields of pure and applied science, and experiments have been made by a large number of workers. As the published results are very discordant, an attempt has been made to clear up the causes of some at least of the disagreements between various methods. The following are some of the fields of research in which the problem occurs:

(*a*) The ascent of sap in plants, and the ejection of spores from ferns.
(*b*) The formation of bubbles in liquids, e.g. blood supersaturated with gas.
(*c*) The design of ships' propellors.
(*d*) The design of echo-sounding and similar equipment, and the study of the biological effects of supersonic waves.
(*e*) The study of surface phenomena attending an under-water explosion.

It is believed (Dixon and Joly, 1895) that the ascent of sap is due, at least in part, to actual hydrostatic tension which is supplied by capillarity in the passages of the leaf. The ejection of spores (King, 1944) is sometimes achieved by a catapult-like mechanism depending on the rupture of a gradually-drying drop of water. Near ships' propellers, negative pressures occur due to the inertia of the water; near sources of high-frequency sound, the amplitude may exceed the hydrostatic pressure so that the water is in tension for part of the cycle. Negative pressures are not the direct result of an under-water explosion, but occur when the resulting pressure-pulse is reflected at a yielding surface such as the free surface of water. In order to keep the pressure at such a surface zero or small, the pressure-pulse is reflected in part as a tension-pulse, so that all the water between the charge and the free surface is subjected first to pressure, then to tension.

§ 2. SUMMARY OF RESULTS OBTAINED BY OTHER WORKERS

Useful reviews of the literature have been given by Harvey and others (1944) and by Vincent (1941; 1943 a and b), who have themselves made valuable contributions to the subject. For convenience we subdivide the results into three classes, according to the tensions estimated. (We have been unable to examine all the papers ourselves.)

Class A. Maximum tensions over 100 *atmospheres*

The chief work in this class is that of Dixon (1909). His method was that introduced by Berthelot of heating liquid in a sealed tube until it filled, then cooling the whole until the liquid broke. Values of tension were then estimated from the corresponding temperature difference, assuming values for compressibility of water and coefficient of expansion determined at zero pressure, due corrections being made for the thermal expansion of the glass and its elastic deformation. He estimated tensions up to 200 atmospheres.

To this work should perhaps be added that of Kenrick, Wismer and Wyatt (1924), who saturated liquids with gas at 100 atmopheres and reduced the pressure to 1 atmosphere without effervescence, and that of Kenrick, Gilbert and Wismer (1924), who heated water to 270° C. under atmospheric pressure before it exploded. A liquid superheated or supersaturated with gas stands in much the same relation to a liquid under hydrostatic tension as does a solid vessel stressed by pumping air into it to the same vessel stressed by evacuating the region outside. The presence or absence of gas is unlikely to affect the tensile strength of a solid, but might well affect that of a liquid, and might alter the "microscopic" distribution of stress.

Harvey (1944) found that water previously compressed to 1000 atmospheres can withstand tensions (occurring as the result of a hammer blow) estimated at from 100 to 1000 atmospheres, but no definite figures are available. King (1944), from a study of the catapult-like mechanism of the projection of spores in the common fern, arrives at a value of 200 atmospheres for the critical strength of water in the presence of plant materials, but his data would be consistent with much lower or higher values.

Class B. Maximum tensions between 10 *and* 100 *atmospheres*

Near the top of this class stands the work of Budgett (1912), who measured the force necessary to pull apart slightly wet optically flat steel surfaces which had been "wrung" together. On the crude assumption that the liquid film is continuous over the surface, one deduces a value of only 4 atmospheres for the critical tension, but Budgett, from examination of the surfaces after separation, and by comparison of the results under vacuum with those under atmospheric pressure, estimated that air could reach most of the inner surfaces and that the effective area of the water was only about 7% of that of the steel. On this basis he estimated that the critical tension water could stand was about 60 atmospheres. He gave arguments which suggest that only a little of this tension is due to capillarity, and also that the failure occurs in the liquid rather than at the steel surfaces. Perfectly dry surfaces did not adhere, so the attractions of the steel surfaces contributed nothing to the effect.

Next come a number of measurements by Worthington (1892) and Meyer (1911). Their method is similar in principle to that of Berthelot, that is, the tension is produced by cooling liquid in a glass vessel, but here the tension is actually measured instead of being inferred from the temperature difference. Worthington measured the tension by noting the change of volume of an ellipsoidal bulb containing mercury (and enclosed in the apparatus), Meyer by means of the rotation of a glass spiral containing the liquid. These devices had to be calibrated with positive pressure and extrapolated to the region of tension. Meyer obtained a value of 34 atmospheres for water and Worthington one of 17 atmospheres for ethyl alcohol.

Yet another modification of Berthelot's method was introduced by Vincent (1943 b). In order to make sure that the pressure in the tube was truly zero at the moment of filling, the " sealing " was carried out by freezing part of the liquid in a capillary tube attached to the main tube and open to the atmosphere at the far end. Unfortunately the liquid used was mineral oil and not water, but the mineral oil had been previously tested by the ordinary Berthelot method and had given a value of 119 atmospheres (maximum) for the tension compared with 157 atmospheres for water under similar conditions. By the modified Berthelot method, in which the pressure at the instant of sealing was definitely known to be atmospheric, values for the critical tension ranging from zero up to 25 atmospheres were obtained. In a further set of readings, where the tube was sealed, the liquid heated to generate pressure, the tube unsealed and the pressure adjusted, and the tube sealed again and cooled in order to apply tension, values for the critical tension varying from 0·9 to 24·5 atmospheres were obtained, which appeared to indicate that the pre-application of pressures of the order of 10 to 100 atmospheres did not alter the results in order of magnitude.

Momentary tensions of the order of 15 atmospheres have actually been measured by piezo-electric gauges in the water near an explosion.

Class C. Maximum tensions below 10 *atmospheres*

Near the top of this class comes the work of Vincent (1943 a). His method consisted of cooling liquid in a vessel, thus sucking it backwards through a capillary tube. (He calls this the "viscosity tonometer" method). By comparison of the rates of flow just before and just after the break one can infer the critical tension from the ratio of these rates of flow, because the pressure in a bubble in the liquid must be just its vapour pressure. Provided the flow is not turbulent, the rate of flow is proportional to the pressure difference. Vincent obtained a maximum value of 7·8 atmospheres for the same oil that he investigated by the Berthelot and modified Berthelot methods. (This method is unsuitable for water because of its low viscosity.) A method of applying tension to a liquid by centrifugal force was introduced by Reynolds (1878), who rotated a sealed U-tube on a lathe, one arm of the tube being full of water and the other containing water and water-vapour. He obtained a value of 4·8 atmospheres for water, and Worthington (1892) applied the same method to other liquids, obtaining a value of 7·9 atmospheres for ethyl alcohol.

Dixon and Joly (1895) obtained a few results, the largest being 7·5 atmospheres, by experiments similar to those of Dixon (1909), but the volume of the cavity formed by contraction of the liquid after failure was obtained by measurement of its geometrical dimensions rather than by temperature difference.

In another series of experiments, Vincent (1941) subjected liquid to tension by the very direct method of enclosing it within a metal bellows and applying tension to the bellows. He obtained maximum values of tension of 2·38 atmospheres for ethyl alcohol and 2·94 atmospheres for the mineral oil.

Values of tension for which cavitation occurs near ship's propellers do not seem to have been measured, and data on quartz or magnetostriction oscillators are hard to find in the literature. Dean (1944) quotes one case where cavitation occurred near an oscillator even though the amplitude of the oscillations of pressure at the surface was only 0·8 atmosphere, and the pressure therefore always positive, according to classical hydrodynamics. He suggested that there might be a further lowering of pressure due to rotational motion. According to the data given by Kornfeld and Suvarov (1944), the amplitude of pressure variation when strong cavitation set in would come out classically at just over 1 atmosphere, but again one cannot exclude the possibility of vortices occurring. It seems unlikely that the tensions are very large.

§ 3. DISCUSSION OF THESE RESULTS

A system consisting of liquid under tension is unstable with respect to a mixture of the liquid and its vapour, so a certain amount of spread in the observed values of tension would not be at all surprising, but the observed spread is enormous. For example, the same observer (Vincent), working with the same liquid (mineral oil), obtained four different values by four different methods, ranging from 2·9 to 119 atmospheres, and a spread of similar extent is found by comparing the results for water. In addition, all the observed results are enormously below the theoretical values. For water the cohesion, or "intrinsic pressure", calculated on the Laplace theory of capillarity, comes out at about 10,000 atmospheres. More refined considerations, taking account of the "hole" structure of the liquid explicitly (Fürth, 1941), still lead to a value of about 3000 atmospheres. A similar discrepancy between the observed and calculated strengths of solids has long been known. Joffé (1928) has shown that part at least of the latter discrepancy is due to surface cracks leading to a non-uniform stressing. Such explanations are not possible for liquids, since stresses cannot be "locked up" in part of a liquid, but must distribute themselves uniformly through it.

It seems certain that the assumptions underlying some of the methods are faulty and that not all of the observers are really measuring the same thing. For example, the liquid may really be failing by tearing away from the supporting surface of the containing vessel so that the properties of this are important. (This certainly seems to occur with the Berthelot method.) This might help to explain some of the discrepancies between the various methods, but hardly that between theory and experiment. (If it was a fact that the cohesion was immensely stronger than the adhesion, the liquid would not wet the surface at all.) Again, it is quite likely that the critical tension in a liquid may depend on whether the stress is applied statically or dynamically. Such effects are familiar in solids. Yet a third possibility is that a prior application of pressure may alter the properties of the liquid under tension or may make it adhere better. Harvey's work (1944) seems to point to an effect of this kind.

We shall now enumerate and discuss some assumptions made by the observers. There is first the assumption that certain physical quantities measured at zero or positive pressure can be assumed to retain the same, or nearly the same, values under conditions of tension. In Dixon's method (1909) it is assumed that the compressibility and coefficient of expansion of water are unaffected in order of magnitude by tension. In Worthington's (1892) and Meyer's (1911) method it is assumed that a calibration of their measuring instruments for positive pressures holds good for tensions. In Vincent's tonometer method (1943 a) it is assumed that the viscosity is unaffected by tension. It seems most unlikely that Worthington's and Meyer's assumption that the changes of shape of a glass vessel are equal and opposite for pressure and tension can be seriously in error, because glass obeys Hooke's law until near its breaking point. There is rather more doubt about the physical properties of liquids under tension, because they have not been measured. If it were possible to apply sufficient tension, the properties of the liquids would certainly be changed, just as they are at high pressures, but it seems certain that these tensions would have to be comparable with the *theoretical* breaking strength—that is, with the energy necessary to remove the molecules from one another's influence, which quantity can be inferred from the latent heat or surface tension. The only experimental results that both involve this assumption and point to tensions at all comparable with the theoretical strength seem to be those of Dixon (1909). Any criticism of Dixon's work on these lines, implying that the extensibility of water may be much greater than the compressibility, and his inferred tensions therefore too high, tends to stultify itself, because it implies that the actual values of tension are lower, and thus lessens the probable departure of the properties of water from those measured at zero pressure.

A second assumption implicit in Dixon's work is that the pressure in the tube is nearly zero at the moment when it fills with liquid. Vincent and Simmonds (1943) have brought forward some evidence that this assumption is in error for his mineral oil, and later in this paper we shall bring forward other evidence for water (also pointing in this direction). An assumption of a different kind is implicit in the Reynolds centrifugal method. One is inclined rather hastily to assume that the liquid in the tube rotates like a rigid body and that the liquid is being stressed statically. Later in this report we shall examine the situation in detail, when it will appear that the stress is dynamic rather than static, and that conditions are in another respect unfavourable for the development of large tensions, so that the centrifugal method cannot fairly be compared with the truly static methods. In addition, it is almost impossible to avoid vibration completely. The resulting accelerations mean that the *maximum* tension actually applied to the water is greater than the " centrifugal " value.

We feel bound to take exception to one statement by Vincent (1941, p. 128): "The only two methods that are theoretically irreproachable, those of Osborne Reynolds and Budgett, give comparatively low results". We have pointed out above that Budgett's experiments give a rather high result if his own interpretation of them is correct, and we see no reason to question it. Secondly, we hope to show below that Reynolds' method does not measure precisely the same physical quantity as do static methods, so that at least in this sense it is not "theoretically irreproachable".

§4. DESCRIPTION OF WORK CARRIED OUT AND OBJECTIVES AIMED AT

The objectives of the work now to be described were threefold.

(*a*) To try and find some reasons for the wide discordance of observed results.

(*b*) To ascertain something about the probable behaviour of water under impulsive, as distinct from static, loading.

(*c*) To investigate possible reasons for the disagreement between theory and experiment.

It is believed that objective (*a*) has been partly attained in that acceptable explanations have been found for the very high results obtained by Dixon using the Berthelot method and the low ones obtained by the centrifugal method (Reynolds, 1878) and by Vincent with the metal bellows. Objective (*b*) requires further work, but the indications are that water is weaker dynamically than it is statically. Harvey's work should help progress towards objective (*c*). He found that strong centrifuging or pre-compression of water to 1000 atmospheres materially increased its power to withstand tension, probably by the removal of stray nuclei and dissolved gases. The work carried out may be subdivided into three parts which will be described separately. The first part was an examination of the Berthelot method of measuring tension. A modification was made which is believed to be capable of giving more reliable results. The second part was an examination of the Reynolds method. The third part was a brief study of cavitation occurring when water is sucked through a constricted tube.

§5. EXAMINATION OF THE BERTHELOT METHOD OF MEASURING TENSION

Suspicion about this method was first aroused by the fact that attempts to repeat Dixon's experiments in wider tubes (about 5 mm. bore), invariably led to the tube bursting, before it filled, however slowly it was heated. A slightly smaller and stronger tube of 3 mm. bore and 1 mm. wall was then used. These experiments were successful but a large percentage of bursts still occurred. (The bursting strength of such tubing is given by Kaye and Laby as 280 atmospheres, but they recommend a large factor of safety to allow for "locked-in" stress). Attempts were made to repeat the work of Dixon and Joly (1895) by measuring geometrically the size of the cavity formed by shaking together all the small bubbles that appeared when the water broke, but it proved impossible to get sufficient accuracy. (It will be remembered that Dixon and Joly obtained several values by this method, of which the *largest* was only 7·5 atmospheres.)

It was therefore decided to use a weighing method, to determine this volume, analogous to Dumas' method of determining vapour density. An account of a typical experiment follows.

A soda-glass tube of bore 3 mm. and wall thickness 1 mm., and about 15 cm. long, was rinsed with acetone, then with distilled water, and dried. It was then sealed at one end and drawn down to a jet near the other end, annealed in the yellow flame and filled with distilled water by heating and cooling. It was then connected to a filter pump and left at room temperature until no more air bubbles appeared. Those which had appeared were coaxed through the constriction by tapping and by alternately removing the tube from the filter pump and replacing it. With the filter pump still running, the water was driven from the centre of the constriction by a small blow-flame. The heat was then increased and the tube sealed. (The finished tubes were about 10 cm. long, weighed about 5-7 grams and contained about 1 c.c. of water. Some were sealed while open to the atmosphere, but it was found much easier to seal under the pump.) The finished tube was transferred to a beaker of water which was frequently stirred and heated until the bubble in the tube had become fairly small. The rate of heating was then reduced to about $\frac{1}{2}°$ C. per min. Since the time-constant for conductivity of heat through such a tube comes out at about 4 sec., there are unlikely to be any serious temperature differences. It was found possible to determine a "filling temperature" for the tube. If the bath was below the filling temperature, a bubble appeared in the tube almost immediately after it was removed from the bath, and expanded rapidly as the temperature fell. If the bath was above this temperature, the liquid continued to fill the tube for several seconds after removal from the bath, finally failing with a "click" very similar to the cracking of glass. With a little experience it was found that the filling temperature could be determined to within 1° C. and in most cases to better than $\frac{1}{2}°$ C. It was found possible to remove the tube from the bath, watch it for a few seconds and decide whether the filling temperature had been reached or not, and then replace it in the bath before a break occurred. The filling temperature was found be to very reproducible. The temperature at which the last visible bubble disappeared could not be determined very accurately, but it appeared to be about 2° to 5° C. below the filling temperature. The temperature at which the liquid broke was next determined by transferring the tube to a smaller bath (actually a boiling tube), containing water at the same temperature, which was allowed to cool naturally in the laboratory at the rate of $\frac{1}{2}°$ to 1° C. per minute, with occasional stirring. The temperature at which the water broke was immediately read and, in most cases, was reproducible to within $\frac{1}{2}°$ C. (To make sure that the rate of cooling was not excessive, one tube was tested by allowing it to cool down in the much larger bath in which it had been heated, the natural cooling rate of which was about 0°·1 C. per minute. The breaking temperature was twice determined in this way, and the results were indistinguishable from those obtained with the quicker rate of cooling.) The tube was then nicked with a file at the constriction, dried carefully, left in the balance-case to attain a steady state and weighed to 0·1 milligram with a catenary-chain balance. The tube was then broken open, filled with water by heating and cooling in the bath, and removed from it when the bath had attained the breaking temperature. The tube and piece cut off were dried carefully and again weighed to 0·1 milligram. (A "blank" experiment showed that, provided the break was clean, there was no measurable loss of weight due to possible splinters of glass.) The difference, usually of the order of 2–3 milligrams, represents, after correction is made for the yielding of the glass, the contraction of the water after failure. The tube was finally weighed empty to determine the total mass of water. At the suggestion of Professor Rideal, four tubes were tested after having been cleaned with hot chromic acid instead of acetone, and in one of these an approximately ½% soap solution was used instead of water. The results were very similar to the others.

A second series of experiments was carried out on identical lines except that a few steel chips or a turning were included in the tube. The steel was washed with acetone, dried by exposure to air, then rinsed with water. The filling temperature was not so easy to determine as with the plain water and exhibited a decided tendency to rise, settling down to a value 3° to 5° c. above its initial value after a few cycles. The filling temperature of one tube was determined and again determined 10 days later. It was found to have risen 7° c., but no rust was visible on the steel. This anomaly is no doubt due to the absorption of dissolved gases by the steel, thus increasing the temperature necessary to develop a given pressure in the tube. (It was never observed with the plain water tubes.) The breaking temperature was always very reproducible, but the "clicks" seemed to be noticeably feebler than with plain water. The results are given in tables 1 and 2.

Table 1. Results for plain glass tubes

Filling temp. (° c.)	Breaking temp. (° c.)	Extension (%)		Method of sealing tube	Remarks
		Apparent	True		
82	59	1·35	—	Under atmosphere	
80·5	67·5	0·79	0·34	,, ,,	
75	70·5	0·24	0·11	Under pump	
48	33	0·59	0·14	,, ,,	
41	19	0·61	—	Under atmosphere	
59	36·5	1·03	—	,, ,,	
61	38·5	1·08	0·21	,, ,,	
58	40	0·87	0·12	,, ,,	Same tube gave two sets of values
58	37	0·95	0·19	,, ,,	
58	37·5	0·91	0·15	,, ,,	Different tubes
58	40·5	0·85	0·079	,, ,,	
62	48	0·70	0·20	,, ,,	
54·5	40	0·71	0·20	Under pump	Cleaned with chromic acid
63·5	53·5	0·47	0·12	,, ,,	,, ,,
72	59·5	0·70	0·10	,, ,,	,, ,,
61	50	0·55	0·17	,, ,,	Ditto with ½% soap dissolved in the water

Mean true extension=0·16%, corresponding to 32 atmospheres tension.

The columns headed "Apparent extension" were obtained from the filling and breaking temperatures, using the data in the International Critical Tables on the density of water at various temperatures. No corrections have been made for the thermal expansion of glass or for its yield under stress. If we had assumed that the pressure was zero at the filling temperature and had then corrected for the elastic yield, we should obtain tensions comparable with Dixon's (1909).

The columns headed "True extension" were obtained from the weighings described above, in which the percentage contraction of the water after failure was determined. It is necessary to correct these figures for the elastic deformation of the glass, and this was done by the method used by Dixon. The corrected values are those given in the table. These readings do not depend on the filling temperature, and are thus independent of those for apparent extension.

Suppose we have a tension T in the tube, the fractional increase in volume of the water is βT, where β is the coefficient of compressibility of the water. The fractional decrease in the internal volume of the tube is, by Lamé's equation,

Table 2. Results for tubes containing pieces of steel

Filling temp. (° c.)	Breaking temp. (° c.)	Extension (%)		Method of sealing tube	Remarks
		Apparent	True		
72	51	1·09	0·060	Under atmosphere	Steel and brass filings
51·5	40	0·56	—	,, ,,	,, : initial values
58·5	42	0·80	—	,, ,,	,, : values after 10 days
62	51	0·55	0·076	,, ,,	Very small steel chips.
51	43	0·40	0·010	Under pump	Steel turning
72	53	1·01	0·041	Under atmosphere	,, ,,
62·5	48	0·73	0·042	,, ,,	,, ,,
64·5	50	0·74	0·113	,, ,,	,, ,,

Mean true extension =0·058%, corresponding to 12 atmospheres tension.

$$\left(\frac{\Delta V}{V}\right)_{\text{glass}} = \frac{T}{R^2-r^2}\left(\frac{r^2}{k}+\frac{R^2}{n}\right),$$

where

R = external radius of tube, k = compression modulus of glass,
r = internal radius of tube, n = torsion modulus of glass.

Evidently this quantity must be added to βT to obtain the relative change of volume which is given by the weighings. We have

$$\left(\frac{\Delta V}{V}\right)_{\text{obs.}} = \left(\frac{\Delta V}{V}\right)_{\text{water}} + \left(\frac{\Delta V}{V}\right)_{\text{glass}} = \beta T + \frac{T}{R^2-r^2}\left(\frac{r^2}{k}+\frac{R^2}{n}\right). \quad \ldots\ldots(1)$$

$\Delta V_{\text{obs.}}$ and $V_{\text{obs.}}$ are determined directly from the weighings, independently of the temperatures.

We have

$\beta = 49{\cdot}5 \times 10^{-6}$ per atmosphere. (Kaye and Laby.)
$k = 4 \times 10^5$ atmospheres. (Dixon.)
$n = 3 \times 10^5$ atmospheres. (Dixon.)
$R = 0{\cdot}25$ cm. $r = 0{\cdot}15$ cm.

Inserting these values, we obtain the results

$$\Delta V/V_{\text{obs.}} = 56 \times 10^{-6}\,T \text{ (T being measured in atmospheres).}$$

$$\left(\frac{\Delta V}{V}\right)_{\text{water}} = 0{\cdot}87\left(\frac{\Delta V}{V}\right)_{\text{obs.}} = 49{\cdot}5 \times 10^{-6}\,T.$$

$\left(\frac{\Delta V}{V}\right)_{\text{water}}$ is tabulated.

Strictly speaking we ought to allow for the variation of β with temperature. Inspection of Bridgman's data in the International Critical Tables indicates that this variation is fairly slow (actually a decrease with rising temperature), but exact figures cannot be obtained as Bridgman uses a large pressure-interval :—

It will be seen, therefore, that we obtain the following values for tension in water, using the corrected figures for true extension :—

In the presence of glass. Mean 32 atmospheres. Maximum 68 atmospheres.
Minimum 16 atmospheres.

In the presence of steel. Mean 12 atmospheres. Maximum 23 atmospheres.
Minimum 2 atmospheres.

It was thought advisable to obtain an independent check on the validity or otherwise of Dixon's assumption of zero pressure at the filling temperature. One method of doing this was to examine the tube between crossed polaroids, which enabled the changes of stressing of the tube to be followed qualitatively, and a rough determination to be made of the temperature at which the pressure in the tube was zero. A second method, suggested by Sir Geoffrey Taylor, was to enclose the tube in an outer jacket connected to a capillary, which enabled the change in volume of the Berthelot tube when it was cut open, or when it was allowed to cool until the liquid broke, to be determined directly. While it can be stated that the results confirm those of the present investigation, it has been thought advisable to publish the details separately.

§ 6. DESCRIPTION OF MODE OF FAILURE OF WATER

As already stated, the failure occurred with a click. The water remained "crystal clear", even when examined by very strong light, right up to the moment of failure, there being no sign of opalescence such as occurs near the critical point. The first appearance at failure was a cloud of barely visible bubbles, which cloud faded away in about half a second, leaving a small number of bubbles about $\frac{1}{2}$ mm. in diameter. For the tubes with steel in them the cloud invariably occurred near the steel. For the tubes containing plain water it was difficult to say whether the failure occurred in the water or at the glass surface. The tubes were kept standing upright while cooling and they sometimes jumped up a distance of one or two centimetres at failure. The energy associated with this is of the order of 10^4 ergs, a minute fraction of the energy stored up in the liquid under tension. The figure becomes of a more reasonable order of magnitude when one reflects that not all of this energy would go into sound waves and that only the energy directly incident on the bottom of the tube would be available for lifting it, the remainder travelling outwards into the water in the bath.

§ 7. EXAMINATION OF THE REYNOLDS CENTRIFUGAL METHOD OF MEASURING TENSION

After a number of trials, an apparatus was developed consisting essentially of a disc of aluminium rotated by an electric motor. A U-tube was clamped with its long arm along the diameter of the disc and its short arm lying flat on the disc. The long arm was sealed, and the short arm was open to the atmosphere. The speed of rotation was measured in three distinct ways:

(*a*) By tachometer.

(*b*) The motor was really a motor-generator and had two distinct commutators and armature windings. By connecting a voltmeter across one armature and the supply across the other one could measure the voltage developed and obtain the speed once the voltmeter had been calibrated against the tachometer. This calibration was not sensitive to change in voltage of the supply mains.

(*c*) Late in the series of experiments a " Strobotac " (Neon tube with variable frequency of flashing) was kindly loaned by E. in C. Department, Admiralty, which enabled the speed of rotation to be measured very accurately. It was hoped that it would also be possible to observe the cavitation process, but it was too rapid to follow visually.

The disc was 9″ in radius which, after allowing for the clamping arrangements and the length of the column of water in the short arm, gave an effective radius from centre to meniscus of about 8″. The disc could be rotated at speeds of up to 2000 r.p.m., which gave a maximum tension of about 8 atmospheres if it may be assumed that the water moves like a rigid body so that we may apply the ordinary law of centrifugal force. (We shall discuss this assumption later.) Evidently the pressure at the centre is given by

$$p = p_0 - \tfrac{1}{2}\rho\omega^2 r^2,$$

where p_0 is the pressure of the atmosphere, r the effective distance between the centre and the meniscus and ω the angular velocity. Since all these quantities can be measured, the tension at the centre just before break can be calculated. (The velocity of rotation was increased slowly by means of a variable resistance in order to avoid any contribution to the tension from terms involving $d\omega/dt$. It is easily shown that the Bernoulli variation of pressure due to some air being carried round by the disc cannot be large enough to matter.

Attempts were first made to use air-free water, but we found, by a sensitive test to be described later, that, even if water is boiled for many hours in the atmosphere or is left under a vacuum with constant shaking, there are still dissolved gases left. Since the readings for a single sample of water showed wide variations of critical tension it would be difficult to compare the readings on two different days if we used water with uncertain amounts of dissolved gas in it. It was therefore decided to use tap water without treatment. It was found to be saturated, or nearly saturated, with air, as would be the water in many cases in which cavitation occurs in practice. The experiments were carried out with four different bores of tube (1 mm., 2 mm., 3 mm. and 5 mm.). Late in the series it was realized that the distance between the axis of rotation and the sealed end of the tube might also be of significance, so the overall length of the tube was recorded.

§ 8. RESULTS

A summary of the results is given in table 3.

Table 3

Bore of tube (mm.)	No. of experiments	Mean tension (atm.)	Highest reading (atm.)	Lowest reading (atm.)
5	16	0·88	2·2	0·1
3	25	1·23	5·3	Zero
2	24	1·00	5·0	Zero
1	27	2·34	5·65	Zero

For those experiments in which it was recorded, the overall length of tube was plotted against critical tension. It was obvious that the correlation between tension and length is very slight, even if it exists at all. There is, however, clear evidence of an increase of critical tension as the bore of the tube is decreased from 5 mm. to 1 mm., though the results for 3 mm. and 2 mm. show a slight reversal. As we shall see, there are theoretical reasons for expecting such an increase in observed tension with smaller diameter. It will be seen that the highest values are of the same order of magnitude as that obtained by Reynolds (4·8 atmospheres), though we do not know what bore of tube he used.

§ 9. THEORETICAL ANALYSIS OF THE CENTRIFUGAL METHOD

On the assumption that the motion of the liquid in the tube is entirely irrotational, we can discuss it qualitatively by using the well-known equations for motion in an ellipsoidal envelope (Lamb's *Hydrodynamics*, 5th edn. p. 147). This solution implies considerable slipping of the liquid along the boundary near the short axis of the ellipsoid, but it also implies that, in any real liquid, there must be a boundary layer growing with time, and that in time all the liquid in the tube will be rotating with the same angular velocity as the boundary. (We are very grateful to Sir Geoffrey Taylor for pointing this out to us.) The time interval associated with the change-over is easily computed from the equations of diffusion of vorticity, which are of the same form as the equation of heat-conduction. It is of the order of $\eta a^2/\rho$, where η is the viscosity, ρ the density, and a is the radius of the tube. For a tube of 5 mm. bore this is only about 6 seconds, which is short compared with the time taken by an experiment, so the stirring-up of the liquid due to its inertia is not likely to be important.

If however, the angular velocity is not constant, but increases with time, we still get a stirring effect even if the liquid as a whole does not lag behind the rotation of the tube. If we reduce the tube to rest by a transformation of axes, we have to introduce the following field of force to allow for the effect of the rotation

$$F = m\left(\omega\wedge[\omega\wedge\mathbf{r}] + 2\omega\wedge\frac{d\mathbf{r}}{dt} + \frac{d\omega}{dt}\wedge\mathbf{r}\right), \qquad \ldots\ldots(2)$$

where m is the mass of a particular particle of fluid, $\mathbf{r}$ its radius vector, and ω the vector representing angular velocity, which we may suppose to be directed along the z axis. The first term in this expression is the ordinary centrifugal force which produces a pressure-gradient along the tube. The second term is the Coriolis force, which can contribute nothing to the pressure-gradient along the tube unless there is relative motion of the tube and the liquid as a whole. The third term is proportional to the angular acceleration and gives a longitudinal pressure-gradient antisymmetrical about a plane through the z axis and the axis of the tube. The corresponding two-dimensional problem can be solved easily. Suppose we have the liquid moving between two planes at a distance d apart. Let the x axis lie midway between the planes and let the y axis be perpendicular to them. Then we have

$$\frac{\eta}{\rho}\frac{\partial^2 V_x}{\partial y^2} = \frac{1}{\rho}\frac{\partial p}{\partial x} = \frac{d\omega}{dt}\,y, \qquad \ldots\ldots(3)$$

and the solution of this equation for which the tangential velocity vanishes at $y = \pm d/2$ is

$$V_x = \frac{\rho}{6\eta}\left(y^3 - y\frac{d^2}{4}\right)\frac{d\omega}{dt} \qquad \ldots\ldots(4)$$

The maximum velocity occurs at $y = \frac{d}{\sqrt{12}}$ and is given by $V_x = \frac{d^3\rho}{36\eta\sqrt{12}}\frac{d\omega}{dt}$. This implies that if the tube is slowly accelerated from an angular velocity of zero to ω, a particle of liquid may be shifted a distance of the order of magnitude $\frac{\rho\omega d^3}{36\eta\sqrt{12}}$, if d is the diameter of the tube. For our widest tubes d was 5 mm., which makes this distance 13 cm. for a final rate of rotation of 20 revolutions per second, and this is comparable with the total length of the tube. It is therefore quite likely that minute air-bubbles or other nuclei might be carried to the axis of rotation and thus help the liquid to break. (This theory also we owe to Sir Geoffrey Taylor.)

Our observed tensions are in fact larger for the smallest tubes, in agreement both with this and the irrotational theories, but the irrotational theory also predicts a larger observed tension when the axis of rotation is near the sealed end of the tube than when it is near the centre, because of the greater slipping that would occur in the latter case, whereas there is no particular reason for expecting such correlation on Sir Geoffrey Taylor's theory. In fact we looked for correlation between the observed tensions and the distances of the axis of rotation from the sealed end, but we could find no evidence of any.

There is a second mechanism that would tend to bring any minute gas bubbles towards the axis of rotation—the pressure gradient in the tube. As we have seen, the distribution of pressure is given with sufficient accuracy by

$$p = p_0 - \tfrac{1}{2}\rho\omega^2 x^2.$$

The force on a bubble of radius r due to the pressure gradient is given by

$$\frac{4}{3}\pi\rho r^3\frac{dp}{dx}.$$

The force on it due to viscosity is given by

$$6\pi\eta r\frac{dx}{dt}.$$

(Stokes' law is here only a rough approximation owing to the presence of the walls.) The force on the bubble due to acceleration of the surrounding liquid is

$$\frac{2}{3}\pi\rho r^3\frac{d^2x}{dt^2},$$

since the virtual mass is half that of the displaced liquid.

The equation of motion is, therefore,

$$\frac{d^2x}{dt^2}+\frac{9\eta}{2\rho r^2}\frac{dx}{dt}+2\omega^2x=0. \qquad \ldots\ldots(7)$$

Consider a small bubble of radius such that its surface tension just balances the centrifugal tension on the axis of rotation. Assume the values for the constants already taken, i.e. $a=20$ cm., $\omega=40\pi$. The tension at the centre is the 2·15 atmospheres, and the corresponding value of r is $6{\cdot}8\times10^{-5}$ cm., taking the surface tension as 74 dynes/cm. Taking $\eta=0{\cdot}01$ c.g.s. units and inserting numerical values, equation (7) becomes

$$\frac{d^2x}{dt^2}+19{\cdot}4\times10^6\frac{dx}{dt}+3{\cdot}2\times10^4x=0,$$

whence it is clear that the motion of such a bubble would be over-damped and would have a time-constant of the order of 600 seconds.

The conclusion that one reaches from these investigations is that the centrifugal method is theoretically unfavourable for the production of large tensions even if there is no vibration in the apparatus. Any vibration would set up acceleration and reduce the observed tension, because any small bubbles previously existing in the liquid would tend to be brought near the axis of rotation by the lag of the water behind the tube and by the pressure gradient of the centrifugal force, and this process might well take a time comparable with that needed to do the experiment, especially for a very narrow tube. Harvey found that very vigorous centrifuging (acceleration 4000 g) is necessary to remove gas-nuclei from water in a reasonable time. (The accelerations used by us in the above experiments were only about 500 g.)

§ 10. A STUDY OF CAVITATION IN A CONSTRICTED TUBE

A simple apparatus was constructed consisting of a pressure-flask exhausted by a filter-pump, to which were connected a mercury manometer, a stop-cock open to the atmosphere, enabling the pressure to be held at any desired value, and a tube with a constriction connected to the pressure flask by a short piece of rubber tubing. The other end of the tube was immersed in a beaker of the water to be tested. The pressure in the flask was allowed to fall until cavitation just occurred at the constriction. The pressure-difference was then read on the manometer and held constant while the rate of flow of water was measured. A "blank" experiment was then performed with a tube without a constriction, the pressure being adjusted until the rate of flow was the same. At the conclusion of a series of experiments, the tube was cut open at the constriction and its diameter measured by a travelling microscope. The tension can now be calculated if it may be assumed that the flow in the constriction is laminar. If it is turbulent, the actual tension will be greater. An analysis of a typical set of experiments follows:

Bore of tube 0·3 cm.
Mean diameter at constriction 0·078 cm. Area $= 4{\cdot}8 \times 10^{-3}$ cm.^{2}

Tap water.

Cavitates at 50·5 cm. below atmosphere and 4·2 c.c./sec. flow = 875 cm./sec.
Calculated tension, 0·05 *atmosphere.*

Tap water passed once through apparatus.

Cavitates at 58·0 cm. below atmosphere and 4·3 c.c./sec. flow = 895 cm./sec.
Calculated tension, 0·22 *atmosphere.*

Tap water passed 3 *times through apparatus, then left* 3 *hours under vacuum.*

Cavitates at 64·5 cm. below atmosphere and 4·4 c.c./sec. = 920 cm./sec.
Calculated tension, 0·27 *atmosphere.*

The Reynolds number for such rates of flow is of the order of 4000, so the flow may be turbulent, in which case these values are lower limits only. A region of turbulence would obviously be a very unfavourable one for the survival of a metastable state of the liquid, so these values cannot fairly be compared with those obtained under static conditions.

The apparatus was designed with two other objects in mind:

First, to detect the presence of dissolved gases in water. This can easily be done by sucking the water through the constriction until it cavitates, then suddenly stopping the flow by pinching the rubber tube. The pressure rises to atmospheric, the cavitation disappears except for any dissolved gases that may have been released by the cavitation, which remain in the form of bubbles. Tests on water which had been boiled for many hours still showed traces of gases, and the same was true of water left in a vacuum for several hours or water that had passed many times through the apparatus, so it appears to be very difficult to remove nuclei completely from water.

The second object was to try and get a qualitative picture of what probably occurs behind a " breaking-front " in water, e.g. when an explosion pulse is reflected at a free surface, causing cavitation. The direction of flow of water is in general *away* from such a front. For this study a tube about 3 feet long, with a constriction near the far end, was used. For water nearly saturated with air, isolated bubbles appeared a long way beyond the constriction long before they were in evidence near it. When cavitation at the constriction was definitely established, the sizes of bubble just beyond it were distributed almost at random, but the large bubbles moved faster and overtook the smaller ones, so that the distribution showed large bubbles far from the constriction, and small ones near it.

§ 11. DISCUSSION

We think that our examination of Reynolds' centrifugal method gives an acceptable explanation of the low values commonly obtained by it. Our examination of the Berthelot method seems to account for the high readings obtained by Dixon (1909), but it remains to be explained why it needs a high pressure (which from our experiments appears to be of the order of 50–100 atmospheres) to persuade the water to stick to the glass. It was suggested to us by Professor Rideal (private communication) that this might be due to the trapped gas being forced into fissures in the glass. Harvey (1944) has pointed out that if the "advancing" and "receding" contact angles differ, there will, for any given mass of gas in a fissure, be a finite range of pressures for which the size of the gas bubble in the fissure will remain practically constant. However, it does not follow that one would get larger tensions by increasing the intial pressure, because one might alter the shape of the fissures. Vincent's work (Vincent and Simmonds, 1943, p. 378) clearly shows that there is a limit to the excess pressure one can usefully apply, and with our tubes which are near their bursting pressures the limit is probably much lower. One of our tubes was by mistake heated 5° C. above the filling temperature, but the breaking temperature was unaltered. A repetition of the experiment gave a similar result.

If we discount the high readings obtained by Dixon and the low values obtained by the Reynolds centrifugal method we have the following set of values for water in the presence of glass:

16–68 atmospheres (mean 32 atmospheres). (Modified Berthelot.)
34 atmospheres. (Meyer.)

For water in the presence of steel we have:

2–23 atmospheres (mean 12 atmospheres). (Modified Berthelot.)
Up to 60 atmospheres. (Budgett.)

The agreement for the second set of results is still not very good.

There remain the very low values obtained by Vincent, with his tonometer (7·8 atmospheres) and his metal bellows (2–3 atmospheres). Though none of these are for water, they are for liquids which seem to behave like water in that by other methods they have given results comparable with water. The low values with the bellows may well be due to the fact that pressure was not previously applied. Our results show also that low values are quite possible in the presence of steel even *with* the prior application of pressure. The tonometer suffers from one of the defects of the Reynolds method in that there is a gradient of tension which would tend to drive any nuclei towards the regions of greatest tension, and this may account for the low values observed with it.

§ 12. CONCLUSION

Ordinary water, even if it contains dissolved air, can stand static tensions of the order of 40 atmospheres. Under dynamic conditions the value is probably less, and drops to an almost negligible value if the flow becomes turbulent. The work with the Reynolds apparatus establishes that even water nearly saturated with air can stand tensions of up to 6 atmospheres. A tension of 40 atmospheres is ample to pull sap up the tallest trees, and is not inconsistent with the data on the projection of spores.

ACKNOWLEDGMENTS

We wish to thank Sir Geoffrey Taylor for advice and helpful criticism, also the Director of Scientific Research, Admiralty, for permission to publish this work, which was carried out in an Admiralty Establishment.

REFERENCES

Budgett, 1912. "The adhesion of flat surfaces". *Proc. Roy. Soc.*, A, **86**.

Dean, 1944. "The formation of bubbles." *J. Appl. Phys.*, **15**, 446.

Dixon, 1909. *Sci. Proc. Roy. Soc. Dublin*, **12**.

Dixon and **Joly**, 1895. "The ascent of sap." *Phil. Trans.*, **186**, 563.

Fürth, 1941. "On the theory of the liquid state." *Proc. Camb. Phil. Soc.*, **37**, 276.

Harvey *et al.*, 1944. "Bubble formation in animals." *J. Cellular Comparative Physiol.*, **1**, 23.

Joffé, 1928. "Discussion on cohesion." *Trans. Faraday Soc.*, **24**.

Kenrick, **Gilbert** and **Wismer**, 1924. "The superheating of liquids." *J. Phys. Chem.*, **28**, 1297.

Kenrick, **Wismer** and **Wyatt**, 1924. "Supersaturation of gases in liquids." *J. Phys. Chem.*, **28**, 1308.

King, 1944. "The spore discharge mechanism of common ferns." *Proc. Nat. Acad. Sci. Wash.*, **30**, 155.

Kornfeld and **Suvarov**, 1944. "The destructive action of cavitation." *J. Appl. Phys.*, **15**.

Meyer, 1911. *Abh. Deutschen Bunsen Ges.*, **6**.

Reynolds, 1878. *Mem. Manchr. Lit. Phil. Soc.*, **17**, 159.

Vincent, 1941. "Measurement of tension in liquids by means of a metal bellows." *Proc. Phys. Soc.*, **53**, 26 ; 1943 a. "A new method of measuring tension in liquids." *Ibid.*, **55**, 41.

Vincent and **Simmonds**, 1943 b. "Examination of the Berthelot method of measuring tension in liquids." *Proc. Phys. Soc.*, **55**, 376.

Worthington, 1892. *Nature, Lond.*, **42**, 261.

Note added in proof: I no longer consider that the theoretical breaking strength of a liquid is equal to the so-called "intrinsic pressure." It may, in fact, be very much less. I hope to discuss this point in a later paper.

H. N. V. T.

THE BEHAVIOUR OF WATER UNDER HYDROSTATIC TENSION: II

By H. N. V. TEMPERLEY,
King's College, Cambridge

MS. received 26 *January* 1946

ABSTRACT. In continuation of the research described in Part I (Temperley and Chambers, 1946), a further examination of the Berthelot method of producing tensions in liquids has been made. The critical tension of water has been measured by a method which does not assume that the extensibility and compressibility of water are equal, and other evidence in confirmation of this assumption has been obtained. It is concluded that high pressures are necessary in the Berthelot tube to force the final gas bubble to dissolve in a reasonable time, and that this fact is probably due to the low rates of diffusion of gases through liquids. It is also concluded that there is a large discrepancy between the theoretical and the observed strength of water.

§ 1. INTRODUCTION

In Part I of this paper (Temperley and Chambers, 1946) some experiments on the behaviour of water under tension were described. Two methods of producing tension in liquids were examined in detail, the Reynolds centrifugal method and the method due to Berthelot, which consists of sealing water or other liquid in a glass tube together with a trace of air, heating the whole until the liquid expands sufficiently to fill the tube completely, then cooling until the liquid fails by hydrostatic tension and no longer fills the tube completely. This method has been used by a number of observers besides Berthelot, for example Dixon and Joly (1895), Dixon (1909) and Vincent (1943). In Part I the available evidence was reviewed and some experiments by a modification of the Berthelot method were described, and it was concluded that the very high values of critical tension, of the order of 200 atmospheres, sometimes obtained by this method were probably false, and that the true value was probably in the neighbourhood of 30 to 50 atmospheres. It was suggested that the discrepancy was probably due to high pressures being generated in the tubes before they filled with liquid. If this conclusion is correct, it means that the theoretical strength of water is too high by a factor of the order of 100. A discrepancy of similar order has long been known to exist for solids, but the conventional explanations of this in terms of surface-cracks and stress-concentrations seem quite inadmissible for liquids. In view of the possible theoretical importance of the conclusions of Part I, it was thought desirable to obtain independent confirmation of them by at least one new method. This paper describes a further examination of the Berthelot method and some new modifications of it.

§ 2. PRELIMINARY EXPERIMENTS

A number of tubes were made up by the method described in Part I with "filling temperatures" ranging from 30–80° C. (It will be remembered that the "filling temperature", as defined in Part I, was the lowest temperature at which a tube would fill completely and thus enable the water to be subjected to tension on cooling.) These tubes were heated to the filling temperature in a beaker of water which was then allowed to cool slowly. The state of stress in the glass could be followed qualitatively by viewing the tube between crossed pieces of Polaroid, the glass becoming doubly refracting under stress, and it was invariably found that the glass was under stress at the filling temperature, that the stress decreased to a small or zero value as the temperature fell, only to increase again as the breaking temperature was approached, finally vanishing abruptly when the water inside the tube broke under tension. Unfortunately, it was not possible to locate with any accuracy the temperature at which the stress in the tube was zero, there being an uncertainty in either direction of the order of 2 or 3° C., corresponding to about 10 to 20 atmospheres. This photo-elastic method in its present form could, therefore, only give very rough measurements of tension, but it does provide qualitative confirmation of the work described by Temperley and Chambers in Part I, by proving that the pressure in the tube is *not* small at the filling temperature, and that the temperature at which the pressure does vanish is nearer the breaking temperature than the filling temperature.

The actual process of disappearance of the bubble of gas as a tube was heated slowly to the filling temperature was watched by means of a low-power microscope (magnification about 30). It would have been of little use to employ a higher power, because the image of the bubble became blurred as the filling temperature was approached, presumably owing to the stresses destroying the optical uniformity of the glass. It was, however, possible to identify definitely the instant at which the last bubble disappeared, and it was found that, if the tube was immediately removed from the bath, the water subsequently broke under tension, so that the temperature of disappearance of the bubble was identical with the filling temperature. (In the work described in Part I the disappearance of the bubble was judged visually, but without a microscope, and the two temperatures were thought to be separated by a few degrees.) It is hard to see why a minute gas bubble should be able to persist for long periods in the presence of liquid at high pressure, but this appears to be due to the very low numerical values of diffusivities. We shall return to this point in more detail later on.

It was thought possible that something analogous to the cloud-chamber phenomenon might occur with water in its metastable state under tension, in that an energetic charged particle might act as a nucleus and initiate the failure of the liquid. Various tubes have been cooled in the presence of a beta-particle source, but the results have so far been completely negative.

Attempts were also made to modify the Berthelot experiment for water by lowering tubes gradually into brine kept at about -10° c. by a freezing mixture, instead of by heating them. The expansion of the water as it froze set up pressure and compressed the gas bubble remaining in the tube, exactly as in the ordinary version of the experiment, but the freezing had to be done much more cautiously than the heating in order to get the tube to fill without bursting. It was found that water adheres to ice less strongly than it does to glass, in that failure (when the tube was removed from the brine and allowed to warm up naturally in the laboratory) invariably occurred by the formation of a few fairly large bubbles at an ice-water surface and without the sharp click that is so characteristic of the ordinary Berthelot experiment. One might interpret this result as meaning that water cannot exert tension on ice in equilibrium with it. One might expect water to adhere to ice better than to other solids, but the result becomes reasonable when one reflects that, at pressures less than that of the triple point, the ice would sublime into vapour directly, so that the bubbles formed may have come directly from the ice. The experiment, then, may simply mean that ice cannot assume a metastable state under tension in the way that water can. (It is well known that, while water can be supercooled to a considerable degree, ice cannot be superheated.)

§3. QUANTITATIVE MEASUREMENTS

A serious defect of the work described in Part I was that taking a measurement necessarily involved the destruction of the tube when it was cut open in order to ascertain (by a weighing method) the total volume of cavity formed by the liquid when it broke. The reproducibility or otherwise of the critical tension associated with a given tube could then only be inferred indirectly from the reproducibility or otherwise of the breaking temperature. It will be remembered that in the work described in Part I the breaking temperature was found to be very reproducible except for one tube.

The method used was suggested by Sir Geoffrey Taylor, and it is similar in principle to that used by Worthington (1892). The Berthelot tube was surrounded by a water-jacket which ended in a fine capillary which was open to the atmosphere and was furnished with a scale to enable the movement of the meniscus to be followed. An experiment was carried out as follows: The whole was heated just above the filling temperature for the Berthelot tube and was then cooled to a temperature a few degrees above the breaking temperature, the open end of the capillary dipping into a reservoir so that the whole jacket was full of water. The reservoir was then removed and the cooling continued. The meniscus gradually fell in the capillary, and its position was read every half-minute and a cooling curve plotted. At the breaking temperature, the stress on the glass of the Berthelot tube was suddenly released as the tension vanished, and the tube consequently increased in volume, as was shown by a sudden rise in the level of the meniscus, and a consequent discontinuity in the cooling curve. The size of this discontinuity, together with a knowledge of the cross-section of the capillary and the dimensions of the Berthelot tube, enabled the tension inside the Berthelot tube to be estimated.

A slight complication of this method is that the result would be vitiated if there were a large calorimetric effect when the liquid broke, because the heat released would cause the water in the jacket to expand, and this would be superposed on the effect due to the change of volume of the Berthelot tube. Although consideration of the various heat capacities involved showed that the error would not be serious even if all the energy stored up as tension in the water appeared as heat, it was considered desirable to make an estimate of the order of magnitude of the calorimetric effect. This was done by using methylated spirit in the jacket instead of water. Methylated spirit has a smaller heat capacity per unit volume, but a much larger coefficient of expansion than water, thus increasing the contribution of any calorimetric effect to the observed volume change. To make this contribution as large as possible, the tube and jacket were allowed to cool in the laboratory, the outer bath being removed altogether. The rate of fall of the meniscus in the capillary was kept small by using a Berthelot tube whose breaking temperature was only just above room temperature. Such tubes are quite easily made by the technique of sealing-off under a filter-pump described in Part I. A summary of the results obtained with this tube is given in table 1.

Table 1. Summary of results with first tube

Group	No. of expts.	Mean vol. change (10^{-4} c.c.)	Estimated tension (atmospheres)			Liquid in jacket
			Mean	Max.	Min.	
A	8	3·9	18·5	27	13·5	Water
B	11	7·8	37	44·5	25	Water
C	5	9·9	46·5	53·5	31·5	Meth. spirit

Details of apparatus

Berthelot tube: Effective length 10 cm. Internal radius 0·19 cm. External radius 0·29 cm. Filling temperature 47° c. Breaking temperature 18°·5 c. Temperature of zero pressure estimated photoelastically to be $26\cdot5 \pm 2°$ c. Water equivalent 0·55 gm. Internal volume 1·17 c.c. External volume 2·55 c.c.

Jacket: length of wide portion 11·5 cm. Length of capillary 65 cm. Weight 12 gm. Water equivalent 1·9 gm. Volume (available for liquid in jacket with

Berthelot tube in place) 1·77 c.c. Cross-section of portion of capillary used varied between 4·1 and $5{\cdot}9\times10^{-4}$ cm^2.

In group A the volume change was estimated from direct observation of the meniscus; in groups B and C it was deduced from cooling curves. Owing to the small cross section of the capillary, the movement of the meniscus was sometimes sluggish, and it continued to rise for several seconds after the water in the Berthelot tube broke. It is thus not surprising that the results in group A are low. The systematic difference between groups B and C does suggest that there was a calorimetric effect, a hypothesis consistent with the shape of some of the cooling curves. (Group C was brought prematurely to an end by the bursting of the tube.)

The tensions were estimated by means of Lamé's formula (see Love, *Elasticity*).

$$\frac{\Delta V}{V} = \frac{T}{R^2-r^2}\left(\frac{r^2}{k}+\frac{R^2}{n}\right),$$

where k is the bulk modulus of glass (4×10^5 atm.) and n is the rigidity modulus (3×10^5 atm.), these values being the ones used by Dixon (1909). We can estimate the order of magnitude of the calorimetric effect in the following way: suppose that x calories are liberated per c.c. of water in the Berthelot tube. If we assume that this heat is shared between the water in the tube, the tube itself and the liquid in the jacket, the rise in temperature is found to be $0{\cdot}335\,x$ ° c. with water in the jacket and $0{\cdot}456\,x$ ° c. with methylated spirit in the jacket. If, however, we suppose that the glass jacket also shares the heat, the figures become respectively $0{\cdot}213\,x$ and $0{\cdot}256\,x$ ° c. We assume the following constants for water and methylated spirit. Water: coefficient of expansion in neighbourhood of 20° c. 20×10^{-5}. Methylated spirit: coefficient of expansion 12×10^{-4}; density 0·8; specific heat 0·6. On this basis we find, by a simple calculation, that to account for the observed difference between the means of groups B and C we must have

$x=0{\cdot}24$ calories/c.c. if heat is not shared with glass jacket,
$x=0{\cdot}43$ calories/c.c. if heat is shared with glass jacket.

The potential energy associated with a tension of 37 atmospheres is 0·88 calories/c.c., so that we may conclude that between 25 and 50% of this eventually appears as heat. Presumably most of the energy appears initially as sound, but is degraded by repeated reflection at the walls of the tubes. Taking a figure of 40%, we see that the values of tension in group B have to be reduced slightly to allow for the calorimetric effect. The corrected values are: Maximum 42 atmospheres; minimum 24 atmospheres; mean 35 atmospheres. These should be compared with the value of 32 ± 10 atmospheres deduced from the photoelastic measurement on the same tube (making the conventional assumption that compressibility and extensibility are equal) and with the values of 16 to 68 atmospheres (mean 34 atmospheres) recorded in Part I, which also involve this assumption, but were obtained from twelve different tubes. The method just described does not involve any such assumption, so that it seems that we have obtained an indirect check on this assumption, but there is still the uncertainty arising from the fact that the check is made by comparing the results from different tubes.

Quite by chance, one of the tubes made during this investigation was found to have a very variable breaking temperature, which provided an opportunity to decide between this assumption and the alternative hypothesis that microscopic bubbles are formed as soon as tension sets in, thus increasing the apparent extensibility. The tube was fitted with a water jacket, and the volume changes at "break" were determined exactly as before, except that this time the tube and jacket were enclosed in a beaker of water. The whole was allowed to cool slowly, and the temperature of the bath at break was recorded. A blank experiment showed that the time-lag between jacket and bath was quite small, and a simple calculation shows that the time-constant for heat conduction through the walls of the Berthelot tube is also of the order of a few seconds only. Although the results are obviously not very accurate, they do seem to point very definitely to the compressibility and extensibility being equal, and they are accordingly recorded in table 2. Worthington (1892) reached the same conclusion for ethyl alcohol.

Table 2. Results for tube with variable breaking temperature

Breaking temperature	Number of experiments	$10^4\times$ observed volume change (c.c.)	$10^4\times$ calculated volume change (c.c.)	Estimated tension (atmospheres)
38	1	4·0	5·5	19
40	2	4·15	4·3	19·5
40·5	1	4·1	4·05	19·5
41·5	2	4·0	3·5	19
43	4	2·4	2·6	11·5
44·5	1	2·6	1·7	12·3

Details of second tube

Effective length 10 cm. Internal radius 0·19 cm. External radius 0·29 cm. Filling temperature 58° C. Breaking temperatures between 44°·5 and 38° C Temperature of zero pressure found photoelastically to be between 45° and 50° C.

The figures headed "observed volume change" refer to the change of the external volume of the Berthelot tube, which was determined by the method already described. The figures headed "calculated volume change" were calculated on the basis that the temperature of zero pressure was 47°·5 C. and that the compressibility and coefficient of expansion of water are the same under tension as at zero pressure. The change in volume was calculated by Lamé's formula as before. The agreement between calculated and observed values is only rough, but does seem definitely to rule out the possibility of any big change in the compressibility of water when tension is applied. As the method is obviously not capable of giving accurate results, an extended series of observations was not considered worth while.

§4. DISCUSSION

We seem to have established that it is necessary to develop pressures of the order of 50 to 100 atmospheres inside the Berthelot tubes in order to cause the final bubble of gas to disappear in a reasonable time, and that this fact has vitiated some of the earlier measurements. We have now to find some explanation of this fact. We consider the diffusion of gas from the neighbourhood of the bubble. For simplicity, we assume spherical symmetry, and neglect any effect of the glass walls. The equation governing the concentration changes is

$$\frac{\partial c}{\partial t}=n\left(\frac{\partial^2 c}{\partial r^2}+\frac{2}{r}\frac{\partial c}{\partial r}\right). \qquad \ldots\ldots(1)$$

The *International Critical Tables* give n as of the order of 10^{-5} for all common gases (measured in c.g.s. units). We take $c=1$ to mean saturation, and assume boundary conditions $c=1$, $\frac{\partial c}{\partial r}=0$ at $r=a$, the surface of the bubble, where a is a slowly varying function of time. To obtain an estimate of the lifetime of a bubble of radius a, we may, to a first approximation, assume a to be constant, and calculate the rates of decay with time of those solutions of equation (1) which satisfy the boundary conditions initially. This will give us a measure of the rate at which gas diffuses away from the neighbourhood of the bubble. The relevant solutions of equation (1) are of two types,

$$S_k=\frac{\sin kr}{r}\exp(-nk^2t), \quad C_l=\frac{\cos lr}{r}\exp(-nl^2t). \qquad \ldots\ldots(2)$$

In order that the derivatives of these solutions should vanish at $r=a$, we must have $ka=\tan ka$ and $la=-\cot la$. The solution which decays with time the least rapidly is given by $la\sim 2$, so that our estimate of the time-constant for the diffusion of gas away from a bubble of radius a is $\frac{10^5 a}{4}$ seconds if a is measured in centimetres. To reduce this time-constant to a reasonable value, say 10 seconds, it follows that we must compress the gas-bubble to a radius of a few tenths of a millimetre, so that the inferred pressures of the order of 50 atmospheres needed for the tubes to fill in a reasonable time are quite reasonable.

A microscopic examination of the process of filling of tubes did show that, when more than one bubble was initially present, the small ones disappeared long before the large ones, and also that, if the bubble was confined in the narrow neck of the tube, diffusion thus being hindered, the tube was much more difficult to fill than if the bubble was shaken into the wider portion. Investigation at temperatures slightly below the filling temperature indicated that the bubbles did still diminish in size, although very slowly, so that reduction of the filling temperatures might be achieved at the expense of prolonging the time of the experiments. The above numerical estimate of the time-constant of the bubbles takes no account of any possible effect of pressure on diffusivity, but increase of pressure is not likely to *increase* the diffusivity. The author has been unable to find any numerical data on this point.

§ 5. CONCLUSIONS

We therefore seem to be forced to the conclusion that the strength of water in the presence of glass is of the order of 30 to 50 atmospheres, and that the higher values reported by Dixon are due to the high pressures required to make the tubes fill in a reasonable time, which may, in turn, be traced to the low numerical values of diffusivities in liquids. The discrepancy between the theoretical and the observed strength of water seems to be confirmed, and to merit further theoretical investigation.

ACKNOWLEDGMENTS

I wish to thank Sir Geoffrey Taylor for suggesting this investigation; also Sir Lawrence Bragg for giving me facilities for it. I also wish to thank the Leverhulme Trust for a research grant.

REFERENCES

DIXON and JOLY, 1895. *Phil. Trans.*, **186**, 563.
DIXON, 1909. *Sci. Proc. Roy. Soc. Dublin*, **12**.
TEMPERLEY and CHAMBERS, 1946. *Proc. Phys. Soc.*, **58**, 420.
VINCENT, 1943. *Proc. Phys. Soc.*, **55**, 376.
WORTHINGTON, 1892. *Nature, Lond.*, **42**, 261; *Phil. Trans.*, **183**, 355.

THE VARIATION OF RESOLUTION WITH VOLTAGE IN THE MAGNETIC ELECTRON MICROSCOPE*

BY V. E. COSSLETT,
Electrical Laboratory, Oxford

MS. received 4 *March* 1946

ABSTRACT. The spherical and chromatic aberrations, the diffraction error and the total aberration of a magnetic electron lens are discussed, as regards variation with voltage, on the basis of theoretical data for the dependence of the aberration coefficients and focal length on lens power. The total aberration shows a minimum value at a given value of the lens power, in a lens of fixed aperture and maximum field strength. The minimum does not occur at either the minimum focal length or the minimum of the spherical aberration coefficient, but at intermediate values of these quantities. When the maximum field is 10,000 oersteds, the minimum resolution is 10–12 A. at an accelerating voltage of 50 kv., the corresponding focal length being 2 mm. If the aperture is adjusted, in a lens of otherwise fixed dimensions, to the optimum value at each voltage, the total aberration varies very little at high voltages. If the lens dimensions are appropriately increased as well as the aperture, the total aberration can be progressively reduced with increasing voltage.

§1. INTRODUCTION

ELECTRON lenses are subject to the same five aberrations as are optical lenses: astigmatism, coma, distortion, curvature of the field and spherical aberration. They also show chromatic aberration when the incident beam is not monochromatic. Magnetic electron lenses show three additional errors, arising from the rotation of the image in the focusing field. In light microscopy the correction of lenses is so far advanced that the primary limitation on resolving power is the diffraction effect, which depends directly on the wave-length and inversely on the aperture; large apertures are therefore employed. Magnetic lenses, on the other hand, are still in a primitive state of development and are almost entirely uncorrected. The form which has been evolved for electron microscopy, largely by trial and error, suffers from a high degree of spherical aberration, which depends on the third power of the aperture; this is the limiting factor on resolution, the other aberrations being at present negligible in comparison. It is therefore necessary to operate magnetic lenses with such minute apertures (10^{-2} to 10^{-3} radian) that the diffraction effect, negligible at large and intermediate apertures on account of the small wave-lengths of the electrons, becomes important and a working compromise has to be reached: it has been usual to employ that aperture size for which the spherical and diffraction errors are equal. In order to keep the chromatic aberration (proportional to the aperture) smaller than these errors, it is then necessary to provide a high degree of smoothing of the voltage supply.

---

* The chief conclusions of this paper were briefly reported to the Electron Microscope Conference held in Manchester on 16 January 1946. Comments have now been added on the treatment of the problem in D. Gabor's *The Electron Microscope*, which has since become available.

The effect of variation in accelerating voltage on the resolution limit is complex. The diffraction error depends on the equivalent wave-length of the electron beam, and hence inversely on the root of the voltage; it is therefore *prima facie* advantageous to operate at high voltage. As regards spherical aberration, previous discussion (on lines already summarized (Cosslett, 1945)) has led to the conclusion that it can be continuously decreased by raising the accelerating voltage and by reducing the lens dimensions. The former conclusion rested on the application to strong lenses of an approximate relation for the aberration of weak lenses (Scherzer, 1933), where the spherical coefficient depends inversely on the voltage. A reduction in dimensions appears to be advantageous because the degree of spherical aberration (at given voltage and lens field strength) depends on the product of the focal length of the lens and this coefficient, which is dimensionless and characteristic of the shape of the field distribution. The focal length falls in proportion to the lens dimensions, having a minimum value in a given lens equal to the half width (a) of the distribution curve at half the maximum field strength (H_0). For the purpose of evaluating the theoretical resolving power of an electron microscope, the value of the product (coefficient × focal length $= C_s$) has sometimes been taken as unity, but more usually a minimum value for a given field has been calculated or estimated, under the assumption that optimum resolution occurred at minimum C_s, and that this minimum value could be attained in practice.

It seemed desirable, from such a treatment of the problem, to employ the maximum practicable voltage in electron microscopy, in order to achieve reduction in both diffraction and spherical errors. However, attempts to improve resolution by operating magnetic microscopes at very high voltages (200-300 kv.) did not yield the expected result, either in the U.S.A. or in Germany. There has been some discussion (Ardenne, 1942; Dosse, 1942) on the reasons for this failure, but with no agreed conclusion. Gabor (1945) has also given attention to the problem, but underestimates the effect of voltage variation, owing to his use of the weak lens approximation for the dependence of focal length on voltage, and overestimates the effect of varying the lens dimensions, owing to a misinterpretation of Ramberg's results. The present investigation shows the determining factors to be the variation of the spherical aberration coefficient with voltage, and the saturation limit to the field strength set by the properties of the iron in the poles of the lenses.

An electron lens differs from its optical counterpart in being a "dynamical" lens, its power being continuously variable by changes in accelerating voltage V and field strength H, as well as in its dimensions. Its aberrations are complicated functions of V, H and their derivatives with respect to axial distance, as well as of the aperture. In a lens of given dimensions and, therefore, of given field distribution, both the aberration coefficient and the effective focal length will depend upon its power, which is represented by an expression of the form $cd^2H_0{}^2/V$, where c is a constant and d is a characteristic dimension, which may be the diameter D of the pole pieces or the field-width parameter, a. In practice it is desirable to operate with the maximum possible value of the field strength at the centre of the lens (H_0), as determined by the saturation properties of the iron circuit, in order to employ maximum voltage at a given lens power and thus minimize the diffraction error. Primary importance, therefore, attaches to the variation of the characteristics of a lens with voltage, at constant H_0.

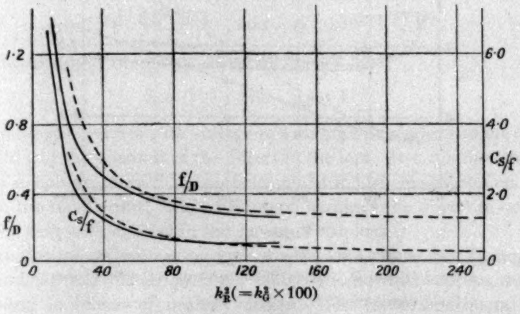

Figure 1.

For Glaser's field distribution, $H_z=H_0/(1+z^2/a^2)$: $\left(k_G^2=\dfrac{ea^2H_0{}^2}{8mV}\right)$.

For Ramberg's field distribution, $H_z=H_0 \operatorname{sech}^2 az$: $\left(k_R^2=\dfrac{D^2H_0{}^2}{2{\cdot}63^2V}=100k_G^2\right)$.

Sufficient data are now available on the theoretical behaviour of lens aberrations with lens power (Glaser, 1941; Ramberg, 1942) to permit calculations of lens performance which take into account all the variables. It proves to be the case that the spherical aberration coefficient (C_s/f) falls with increasing lens power (figure 1), but at constant H_0 it must therefore increase with accelerating voltage, since power $\propto 1/V$. This increase is so rapid as to offset the parallel reduction in the diffraction error with voltage. The total aberration (assuming the chromatic error to be negligible) thus exhibits a flat minimum in relation to voltage (figure 2), for a given value of H_0, and this minimum does not occur at the minimum of either the focal length or the spherical aberration coefficient of the lens, as assumed by Dosse, Ardenne and others. If the magnetic field is increased, the minimum aberration occurs always at the same numerical value of the lens power; its absolute value decreases if the voltage is correspondingly raised, maintaining the power of the lens constant at this optimum value. On the other hand, it now appears that reduction in lens dimensions offsets the favourable effect of voltage increase. Once the saturation field strength has been reached, reduction in dimensions demands a lower voltage for optimum resolution, a conclusion arrived at experimentally by Dosse (1941). Conversely, attempts to raise the working voltage with lenses of fixed dimensions and limited field strength can only result in a deterioration in resolution.

Once the saturation field strength has been reached, it is now shown that further improvement in resolution can only be attained by increasing simultaneously both the working voltage and the lens dimensions. The resolution improves in inverse proportion to the fourth root of the voltage if the dimensions are increased in proportion to the root of the voltage. On this basis it is possible to design a magnetic electron microscope having a theoretical resolving power of a few Ångström units only, together with the high penetrating power required for the examination of many specimens, particularly of a biological nature.

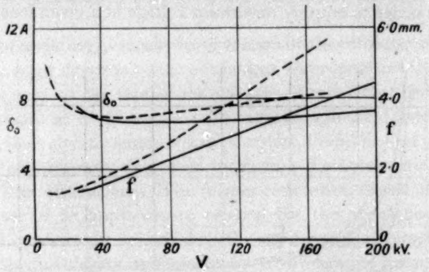

Figure 2.

Focal length, and minimum error ($\delta_0=\delta_S=\delta_D$), as functions of accelerating voltage at fixed H_0 and D. ($H_0=10{,}000$ oersteds ; $D=4a=5$ mm.).

Field distribution, $H_z=H_0/(1+z^2/a^2)$: - - - - - - -

,, ,, $H_z=H_0\,\mathrm{sech}^2\,az$: ————

The variation of chromatic aberration with lens power and voltage is also discussed. It shows a flat minimum value for a lens with fixed aperture size. If the aperture is varied to its optimum value for each lens power, as required by consideration of the spherical and diffraction errors, a very flat minimum appears, but at a power corresponding to a higher voltage at fixed field strength.

§ 2. SPHERICAL ABERRATION AND DIFFRACTION

The confusion in an electron microscope image is due primarily to diffraction and spherical aberration. The former effect depends inversely on the value of the aperture (α) and the latter directly on its third power. If δ_d and δ_s are the diameters of the respective discs of confusion, referred back to the object plane, then

$$\delta_d=K\lambda/\alpha \qquad \ldots\ldots(1\,a)$$

and

$$\delta_s=C_s\alpha^3, \qquad \ldots\ldots(1\,b)$$

where λ is the wave-length of the " illuminating " electron beam and C_s is a spherical aberration coefficient characteristic of a given lens. K is a numerical constant dependent on the assumptions made as to when two diffraction discs become resolvable by the eye; we shall assume for the time being that it is unity, and discuss later the results of a more exact wave mechanical treatment (Glaser, 1943). Since the two errors vary in opposite senses with aperture, there is an optimum value of aperture giving minimum aberration. It is usual to assume that this occurs when the two errors are equal (Rebsch, 1938), so that

$$\delta_d=\delta_s=\delta_0=\sqrt[4]{C_s\lambda^3} \quad \text{and} \quad \alpha_0=\sqrt{\lambda/C_s} \qquad \ldots\ldots(2)$$

if $K=1$. We prefer to use δ_0 here, for the optimum aberration at given λ, and to reserve $\delta_{\text{min.}}$ to designate its minimum value as λ is varied by changing the accelerating voltage, V, in a given lens.

From de Broglie's relation, and neglecting relativistic corrections, we have $\lambda=\sqrt{150/V}$ A., and, therefore, with C_s also in Ångström units,

$$\delta_0=150^{\frac{3}{8}}\,.\,\sqrt[4]{C_sV^{-\frac{3}{2}}} \quad \text{and} \quad \alpha_0=150^{\frac{1}{4}}/\sqrt[4]{C_sV^{\frac{1}{2}}}. \qquad \ldots\ldots(3)$$

If C_s were independent of voltage, then we should have

$$\delta_0 \propto V^{-\frac{3}{8}} \quad \text{and} \quad \alpha_0 \propto V^{-\frac{1}{4}}.$$

In fact, the value of C_s varies with voltage in a manner dependent on the form of the magnetic field and the position of the object in the lens. In a microscope objective the object is deeply immersed in the field, and for high magnification should be very close to the focal point. This discussion assumes the object to occupy such a position, varying with change in the power of the lens.

The variation of spherical aberration with lens power has been calculated by both Ramberg and Glaser for field distributions which approximate closely to those found in lenses of present design. The important quantity is C_s/f, which is dimensionless and characteristic of a particular form of field distribution, being invariant with respect to changes in lens dimensions providing that all are changed to scale. Glaser has calculated C_s/f for the field distribution $H_z=H_0\Big/\left(1+\dfrac{z^2}{a^2}\right)$, where H_z is the axial field strength at the point z, H_0 is the maximum value of the field at the axis, and a is a lens parameter related to the half-width of the field distribution curve at $H_0/2$, and in this particular case equal to it. A rigorous solution of the electron path equation is possible with this field distribution, and an exact calculation of aberrations can be made. The power of the lens depends on the integral of H_z^2 taken along the axis from object to image, and, therefore, Glaser specifies the power of a lens in terms of a parameter $k^2=\dfrac{ea^2H_0^{\,2}}{8mV}$. His results for C_s/f and the relative focal length $f/4a$ are given in figure 1 for the conditions of high magnification specified above.

Ramberg investigated the very similar distribution $H_z=H_0\,\text{sech}^2\,az$ by a ray-tracing and numerical integration method. He found the aberration in terms of a constant $C(=C_s/f)$ and specified the lens power by the quotient $D^2H_0^{\,2}/2{\cdot}63^2V$, where D is the minimum diameter of the pole pieces of the lens. Confusion has arisen in comparing his results with those of Glaser, which appear to give much smaller values of the aberration coefficients. However, direct comparison is only legitimate if the two calculations are referred to lenses of the same dimensions and focal length. To this end we here assume $D=4a=5$ mm., and transform Ramberg's results into the set of parameters employed by Glaser; the results are plotted in figures 1 and 2. The two focal-length curves fit very well at a minimum value of $f/a=1$, and the aberration curves also coincide satsfactorily; the degree of divergence is not more than might be expected from the difference in the field distributions operated upon in the two cases. As regards lens power, we find that the above assumption ($D=4a$) leads to a simple relation between Ramberg's and Glaser's parameters: $D^2H_0{}^2/2{\cdot}63^2V=100\,k^2$, to a close approximation.

Both the aberration coefficient and the focal length curves fall with increasing lens power (k^2); that is, with increase of field strength, decrease of voltage or a degree of both together. The limiting factor in lens operation is the saturation of the pole pieces, and, since diffraction and other considerations make it desirable to use the highest possible voltage, it is usually necessary to work with field strengths close to saturation, in order to keep the lens power high and f and C_s/f small. We assume the saturation field to be 10,000 oersteds, and from this value of H_0 and the assumed value of a (or D), the value of the voltage corresponding to any value of k^2 can be calculated. Figure 1 allows C_s to be evaluated, and hence the value of δ_0 to be found for each voltage. Figure 2 shows the results thus calculated for both Glaser's and Ramberg's field distribution. That is, it gives the optimum aberration that can be obtained with a lens having the assumed dimensions (D) and saturation field (H_0). The two sets of results agree very well, showing a minimum aberration ($\delta_{\text{min.}}$) of 7 A. at an accelerating voltage of 60 kv.; but the minimum is flat and the aberration does not exceed 10 A. over a wide range of voltage. The corresponding focal length given by the data of figure 1 is also shown, indicating that the minimum value occurs at the comparatively low voltage of 11·4 kv.; the optimum f, corresponding to $\delta_{\text{min.}}$, is 2 mm. It is clear

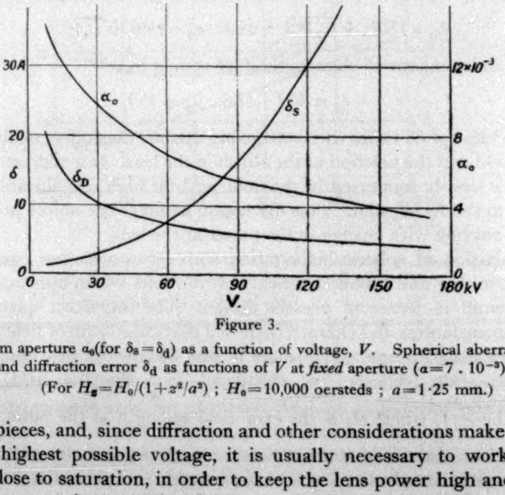

Figure 3.

Optimum aperture a_0(for $\delta_s=\delta_d$) as a function of voltage, V. Spherical aberration δ_s and diffraction error δ_d as functions of V at *fixed* aperture ($a=7\ .\ 10^{-3}$). (For $H_z=H_0/(1+z^2/a^2)$; $H_0=10{,}000$ oersteds ; $a=1{\cdot}25$ mm.)

that minimum aberration does not coincide with minimum focal length nor with minimum C_s, as has sometimes been assumed. The same data yield the optimum value of the aperture for each voltage, from equation (3). The curve (figure 3) shows a regular rise from about $3{\cdot}5$ to $14{\cdot}5 \times 10^{-3}$ with decreasing voltage, having a value of 7×10^{-3} at the voltage giving minimum aberration.

The results show, from the good agreement obtained from calculation on two different approximations to the actual field distribution, that a spherical aberration and diffraction error better than 10 A. is attainable by lenses of the present type, the values of D and H_0 taken being much the same as realized in practice. The combination of these errors with the chromatic aberration to give the total aberration is discussed later, but it may be remarked here that the optimum voltage and aperture found are very close, by accident or design, to those employed in present-day microscopes. The minimum aberration ($\delta_{\text{min.}}$) will always occur for the same value of k^2, as a (or D) and V are varied at constant H_0, but its absolute value will decrease as the voltage is raised. Since $V \propto a^2$ at constant k^2 and H_0, it follows that the limiting value of $\delta_{\text{min.}}$ can only be reduced below 7 A. by increasing both the voltage and the dimensions of the lens. Further consequences of this interrelation are reserved for § 5.

§ 3. CHROMATIC ABERRATION

The diameter δ_c of the disc of confusion (referred to the object plane), due to chromatic aberration alone, is

$$\delta_c = \alpha \, . \, \frac{\Delta V}{V} \, . \, C_c, \qquad \ldots\ldots(4)$$

where $\Delta V/V$ is the fractional fluctuation in voltage and C_c is a coefficient depending on the field distribution and focal length of the lens. The voltage fluctuation is comprised of the residual ripple, $y \, . \, V$, in the high-tension supply to the cathode, together with the thermionic variation in the energy of emission of electrons, x volts. So that

$$\Delta V/V = (x + y \, . \, V)/V = x/V + y. \qquad \ldots\ldots(5)$$

Variation in the focusing field (ΔH) may be included in the factor x, for present purposes.

In a given lens, with constant focusing field H, the chromatic aberration varies with voltage owing to the change of all three variables, α, C_c and $\Delta V/V$. The effect of fluctuations will depend on the relative values of x and y, but only with very high smoothing efficiency (order of $2 \, . \, 10^{-5}$) does the thermionic fluctuation become important, and then only at low voltages. For a given pole-form, and hence field distribution, the non-dimensional ratio C_c/f varies only slowly with lens power, and hence voltage. But the focal length depends on the voltage, at fixed H, and, therefore, the coefficient C_c increases with voltage in the usual range of operation of present lenses, the variation is approximately given by

$$C_c \propto V^{\frac{1}{2}}. \qquad \ldots\ldots(6)$$

The aperture α may be kept constant at a value designed for best average performance over a range of voltages, or it may be varied with the voltage, in which case it will affect the value of the chromatic aberration. Considering first the aperture as fixed, substitution of equations (5) and (6) in equation (4) and differentiating shows that δ_c will have a minimum value when $V=x/y$. Taking Glaser's values for C_c/f for the field distribution $H=H_0\,(1+z^2/a^2)$ and the lens conditions previously assumed ($H_0=10{,}000$, $a=5/4$ mm.), we can calculate the variation of δ_c with voltage at a fixed value of the aperture (figure 4). The aperture chosen was the value (7×10^{-3}), corresponding to minimum spherical and diffraction error ($\delta_{\text{min.}}$) in such a lens. The curve shows a pronounced minimum of about 4 A., but lies below 10 A. over a very wide range of voltage (6–180 kv.). It rises sharply at lower voltages, where the thermionic fluctuation x becomes relatively important. In these calculations the value of the latter was taken as 1 v. and the ripple (y) as 1/40,000. The minimum value of δ_c is then predicted as occurring at $x/y=40$ kv.; the fact that it occurs at about 33 kv. indicates the extent to which equation (6) is approximate.

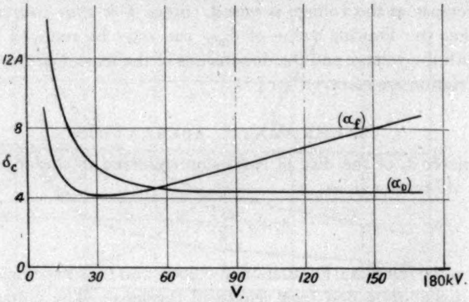

Figure 4.

Chromatic aberration δ_c as function of voltage, at fixed aperture ($\alpha=7\,.\,10^{-3}$), and for variable aperture (α_0). Lens conditions as in figure 3.

In the second place we assume α to be adjusted at each voltage to the value (α_0), giving equality of spherical and diffraction errors, according to figure 3. The latter shows that $\alpha_0 \propto V^{-\frac{1}{4}}$ very closely. Inserting this relation, and equation (6) in equation (4), we have for the chromatic error

$$\delta_c=k\,.\,\Delta V/V=k(x/V+y)$$

from equation (5). Hence δ_c is determined in these circumstances solely by the relative values of the thermionic and ripple fluctuations, and should fall to a constant value dependent on y at high voltages, for which x/V is negligible. The curve calculated from Glaser's values of C_c for this case is given also in figure 4. It has the expected form, the indication of a very flat minimum at high voltage showing that the simplified relations assumed are only first approximations. At 57·3 kv. the two curves for δ_c intersect, this being the voltage for which the value of the aperture assumed for the first curve is equal to α_0 as defined. At lower voltages the equality of δ_s and δ_d requires a rapid rise in α and, therefore, in δ_c, so that the second curve lies above the first. At higher voltages, on the other hand, the value of α_0 falls below 7×10^{-3} and the chromatic error is lower for the second case considered. In both cases δ_c is lower than δ_s and δ_d over the range of voltage from 20 to 150 kv.

§ 4. TOTAL ABERRATION

The values of δ_s, δ_d and δ_c obtained above have to be combined in order to obtain the total aberration observed. It is difficult to form a physical picture of their mutual effect; they are certainly not independent, and, therefore, simple addition is not justified. Both spherical and chromatic aberration will have the effect of producing a number of overlapping diffraction maxima in the image plane. It has been usual to assume, following Ardenne (1938), that they should be superimposed as error functions, and the geometrical mean taken as the total aberration. Such a procedure probably errs on the side of pessimism, but may be provisionally adopted so long as observed aberrations are so much larger than those predicted. We thus have for the total aberration δ_t in the object plane,

$$\delta_t = \sqrt{\delta_s^2 + \delta_d^2 + \delta_c^2} = \sqrt{2\delta_0^2 + \delta_c^2}, \qquad \ldots\ldots(7)$$

when optimum aperture is employed.

The application of this synthesis to the results given above for the individual aberrations yields the curves of figure 5 for the variation in δ_t with voltage both for a fixed aperture and for variable optimum aperture α_0. For the latter condition the total aberration initially falls rapidly with voltage, but then lies between 11 and 13 A. over the range 20 to 170 kv., with a flat minimum at about 50 kv. For a fixed aperture the minimum aberration naturally occurs near the voltage (57·3 kv.) for which α was fixed, as giving minimum δ_0; the increase in δ_s with higher voltage moves the minimum back to about 48 kv. For the same reason the total aberration curve for fixed aperture rapidly rises above that for optimum aperture at high voltages, where $\alpha > \alpha_0$, and, therefore, spherical aberration is large.

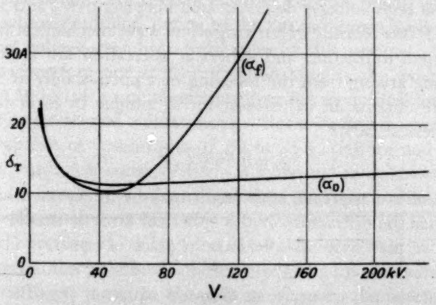

Figure 5.

Total aberration δ_T as function of voltage, at fixed aperture ($\alpha = 7 \,.\, 10^{-3}$), and for variable aperture (α_0). Lens conditions as for figures 3 and 4.

The above procedure has assumed that the combination of spherical and diffraction errors has a minimum when the two are equal. It was on this assumption that equation (2) was derived and equation (3) employed to find the combined aberration when the aperture is α_0, as defined in terms of this equality; so that, neglecting δ_c, we have $\delta_t = \sqrt{2}\delta_0 = 1{\cdot}41\sqrt[4]{C_s\lambda^3}$. But this is only a convenient working rule, having no theoretical basis, and the same is true of the importation of the diffraction expression directly from optics.

Gabor (1945, chapter 7) assumes that the minimum resolution occurs at the true minimum value of the sum of the squares of δ_d and δ_s, found by inserting their analytical relations (equations (1 *a*) and (1 *b*), above) and differentiating. Minimum resolution is then found when $\delta_d = \sqrt{3}\,.\,\delta_s$, whence $\delta_0 = 1{\cdot}33\sqrt[4]{C_s\lambda^3}$. But he also assumes a more favourable value for the diffraction error, justifiable on physiological grounds of density discrimination in the eye: $\delta_d = 0{\cdot}82\,\lambda/\alpha$, in place of relation 1 *a*. His working expressions therefore become, with C_s in cm.,

$$\delta = 1{\cdot}14\sqrt[4]{C_s\lambda^3} = 750\sqrt[4]{C_sV^{-\frac{3}{2}}} \quad \text{and} \quad \alpha_0 = \sqrt[4]{\lambda/\sqrt{3}\,.\,C_s},$$

not sensibly different from the values used in the above calculations (neglecting δ_c):

$$\delta_t = 1{\cdot}41\sqrt[4]{C_s\lambda^3} = 925\sqrt[4]{C_sV^{-\frac{3}{2}}} \quad \text{and} \quad \alpha_0 = \sqrt[4]{\lambda/C_s}.$$

Gabor, however, proceeds to make the weak-lens assumption that *f* is proportional to accelerating voltage and to neglect the variation of C_s/f with voltage. He thus finds $\delta_{\text{min.}} \propto V^{-\frac{1}{4}}$ in a given lens, showing a very slow fall with voltage, in place of the increase at high voltages demonstrated here (figures 2 and 5).

Glaser (1943) has recently given a rigorous wave mechanical treatment of the problem, in which diffraction and spherical aberration are treated together as one phenomenon, arising from the focusing of a portion only of a non-spherical wave-front. He obtains an expression for δ_0 similar in form to equation (2), but with a numerical factor

$$\delta_0 = 0{\cdot}56\sqrt[4]{C_s\lambda^3}. \qquad \ldots\ldots(8)$$

In terms of the above approach, this result may be interpreted in the (extreme) senses that either the diffraction or the spherical error is smaller than assumed. If the former, we may write $\delta = 0{\cdot}42\,\lambda/\alpha$ in place of equation (1 *a*), leaving the spherical aberration equation (1 *b*) unaltered. Assuming minimum error to occur when the two are equal, we arrive at Glaser's equation (8); the practical effect would be that an aperture of $0{\cdot}8\,\alpha_0$ would give a value of δ_0 smaller by 0·56 than that calculated from equation (3). On the other hand, we may write $\delta_s = 1/16\,.\,C_s\alpha^3$, and assume that the diffraction equation remains exactly as equation (1 *a*), in order to obtain equation (8). In this case the practical conclusion is that an aperture of $2\alpha_0$ will give the predicted smaller value of δ_0. If, as is probable, neither of these limiting cases is valid, we shall have numerical factors in equations (1 *a*) and (1 *b*) larger than 0·42 and 1/16 respectively, giving a working aperture rather larger than α_0 for a combined error δ_0 smaller by the factor 0·56.

Glaser himself gives a value for the optimum aperture, $\alpha_0 = 1{\cdot}13\sqrt[4]{\lambda/C_s}$, which yields a diffraction error only 5% smaller than the combined error (equation (8)). This is on his own assumption that the optical diffraction expression applies, $\delta_d = 0{\cdot}6\,\lambda/\alpha$; any other less favourable diffraction expression, such as Gabor's, or equation (1 *a*) of this paper, would lead to a diffraction error greater than Glaser's "total" aberration. Since the spherical aberration must play an appreciable part in determining the total confusion in the electron microscope image, Glaser's conclusions must be regarded as of doubtful application. Their derivation is in any case insecurely based, owing to a lack of accurate information as to the details of the scattering of electrons in the object.

The effect of Glaser's result would be that the theoretical minimum values of the total aberration as here calculated and shown in figure 5 are reduced approximately by one-half. Or alternatively, the same calculated aberration would be obtained with the use of a larger aperture. Since it is not clear exactly how Glaser's result is to be interpreted in its effect on the value of the optimum working aperture, and since other aberrations (including the varying rotation of the image due to voltage fluctuations) have not been taken into account, it has been felt advisable to adhere to the classical method of treatment. The synthesis of aberrations obtained must therefore not be regarded as anything more than an approximate result, erring probably on the high side.

§ 5. DISCUSSION

The results of the present calculations show that there are definite optimum conditions of operation for a given magnetic electron lens. Its dimensions and maximum attainable field strength being fixed, there are particular values of the aperture and operating voltage, and hence of the focal length, which will give minimum total aberration. For present electron microscopes the conditions have been rather fortuitously fulfilled, owing to the flat minimum in the curve of total aberration against voltage even for a fixed value of the aperture. The conditions of operation are therefore not at all critical so long as the working resolution is of the order of 50 to 100 A., much greater than the theoretical limit, on account of limitations in materials and workmanship. But now that research microscopes are becoming capable of resolutions of 20 to 25 A. (Hillier and Baker, 1945), attention to such optimum considerations is more and more essential. The lens data assumed in the above aberration calculations are typical of lenses now in use. Figure 5 shows that their best theoretical performance is 10–11 A. for the total aberration for an aperture of 7×10^{-3} and at a voltage of 50 kv. Practical results are therefore already no more than double those given by theory.*

Attempts have been made in the past to improve the resolving power by operating at very high voltage, up to 300 kv., relying upon a supposed reciprocal relation between voltage and aberration. It is now clear that failure was due to the saturation limit to the focusing field strength, as it was not appreciated that aberrations increase rapidly if the field is not increased in proportion to the root of the voltage—that is, if the power of the lens is not kept constant; this is shown by figure 5. Assuming that the lens is of fixed dimensions, including aperture, and operating at maximum H_0, the total aberration rises rapidly with voltage above a value of about 60 kv. At 172 kv. it is already 67 A., the effective power of the lens having fallen to 0·2 from the optimum value of 0·6.

Since, for a given lens, there is an optimum value of the power (k^2) which gives minimum δ_d and optimum α, it follows from the definition of k^2 that as the working voltage is varied the magnetic field should be varied according to $V^{\frac{1}{2}}$, in order to keep the power constant at this value. From equation (3) it then follows that the value of $\delta_{\text{min.}} \propto H^{-\frac{3}{4}}$. The value of δ_c also falls with increasing field strength and voltage, at optimum focal length, to a degree which depends on whether α is constant or always made equal to the optimum value corresponding to $\delta_d = \delta_s$. The table shows the variation in the several aberrations with field strength for the optimum conditions found above: $k^2 = 0{\cdot}6$, $f = 2{\cdot}1$ mm., when $a = 1{\cdot}25$ mm. $\approx D/4$. As in figure 4, δ_c becomes preponderant at low fields for $\alpha = \alpha_0$, owing to the rapid increase in the latter, whereas for fixed aperture its value is less, whilst δ_d increases, owing to the increase in wave-length of the beam not being offset by an increase in aperture; δ_s is constant in this latter case. It will be seen that the net result gives very much the same total aberration for both fixed and variable aperture at all except very low field strengths. It should be mentioned that this discussion assumes constant high magnification, the focal length being constant throughout at its optimum value.

---

* *Note added in proof*: Hillier (1946) has now reported the achievement in favourable circumstances of a resolution of 10 A.

Table

Variation in individual and total aberrations with field strength H_0 (and corresponding acceleration voltage V) for a lens operated at optimum power ($k_0^2 = 0{\cdot}6$). $f_0 = 2{\cdot}1$ mm.; $a = 1{\cdot}25$ mm.; $H_z = H_0/(1 + z^2/a^2)$. $\delta_t = \sqrt{\delta_s^2 + \delta_d^2 + \delta_c^2} = \sqrt{2\delta_{\text{min.}}^2 + \delta_c^2}$.

H_0 (oers.)	V (kv.)	$\delta_{\text{min.}}$ (A.)	α_0 (rad.)	δ_c (A.)	δ_t (A.)	At fixed aperture, $\alpha = 7 \cdot 10^{-3}$			
						δ_c (A.)	δ_d (A.)	δ_s (A.)	δ_t (A.)
1,000	0·57	40·3	$12{\cdot}7 \times 10^{-3}$	334	339	185	72·5	7·2	199
2,000	2·29	23·9	10·7	73·1	80·5	48·7	36·3	7·2	61·0
4,000	9·17	14·2	9·0	17·9	26·9	14·1	18·1	7·2	23·9
6,000	20·6	10·5	8·1	8·8	17·3	7·7	12·1	7·2	15·8
8,000	36·7	8·5	7·6	5·9	13·3	5·5	8·1	7·2	12·7
10,000	57·3	7·2	7·2	4·5	11·4	4·7	7·2	7·2	10·8
12,000	82·5	6·2	6·8	3·8	9·6	4·0	6·0	7·2	10·0
15,000	129	5·3	6·5	3·1	8·1	3·5	4·8	7·2	9·2

The necessary conditions for design and operation of high-voltage lenses of high resolution follow from the above discussion, and are further considered in a subsequent paper. The expression for the lens power, $k^2 = ea^2H_0^2/8mV$, shows that the power and focal length can be kept constant at their optimum values, as the voltage is raised, by increasing a and H_0. Once the saturation field strength is reached, the remaining variable is a, the field width parameter. Hence, to ensure optimum resolution at high voltages, it is essential to increase the dimensions of the lenses, and not to decrease them, as has been the tendency empirically followed hitherto.

Gabor (1945, chapter 12) has noted, from the results of Ramberg, the fall in C_s/f with increase in lens dimensions at constant H_0 (cf. figure 1). He fails to observe, however, that it is f/D, and not f, which falls also with increasing lens power. If such an increase is brought about by changing the lens dimensions alone, at fixed field and voltage, it proves that f itself remains almost constant, the fall in f/D being roughly proportional to the increase in D, the lens diameter.

Hence only a small improvement in resolution results in such a case, due to the reduction in C_s. On the other hand, the above discussion has shown that better resolution is obtained if the lens power (and hence f/D) is kept constant by increasing both V and the lens dimensions. C_s/f also remains constant, whilst f increases. The increase in $\sqrt[4]{C_s}$, however, is more than offset by the effect of increasing the voltage, since $V^{-\frac{3}{8}}$ enters into the expression for δ_0 (equation (3)). It proves, in fact, that in these circumstances $\delta_0 \propto V^{-\frac{1}{4}}$.

If the maximum value of H_0 is 10,000 oersteds in the lens discussed earlier, then its dimensions will need to be increased some 2·4 times for operation at 300 kv. instead of at 57·3 kv., since at constant k^2 and H_0 we have $a \propto V^{\frac{1}{2}}$. Hence the pole diameter will need to be about $2{\cdot}4 \times 5 = 12$ mm., and the optimum focal length will be approximately 5 mm. For the same magnification the microscope will need to be increased in length by the same factor, since the object distance is effectively the same as the focal length. Such an inconvenient increase in size can be avoided either by working at a lower electronic magnification, and employing subsequent photographic enlargement from fine-grain plates, or by resorting to multi-stage operation, with several lenses each giving small magnification. The latter solution appears to be most satisfactory if high voltage operation is required primarily for the higher resolution it makes possible, and not simply for the greater penetrating power of the beam, allowing thicker specimens to be observed at rather rough resolution.

REFERENCES

ARDENNE, 1938. *Z. Phys.*, **108**, 338 ; 1942. *Ibid.*, **118**, 384.
COSSLETT, 1945. *J. Sci. Instrum.*, **22**, 170.
DOSSE, 1941. *Z. Phys.*, **117**, 422; 1942. *Ibid.*, **118**, 375.
GABOR, 1945. *The Electron Microscope* (Hulton Press).
GLASER, 1941. *Z. Phys.*, **117**, 285; 1943. *Ibid.*, **121**, 647.
HILLIER, 1946. *J. Appl. Phys.* **17**, 307.
HILLIER and BAKER, 1945. *J. Appl. Phys.*, **16**, 469.
RAMBERG, 1942. *J. Appl. Phys.*, **13**, 582.
REBSCH, 1938. *Ann. Phys. Lpz.*, **31**, 551.
SCHERZER, 1933. *Z. Phys.*, **80**, 193.

NEW REGULARITIES IN VIBRATIONAL SPECTRA

By K. M. GUGGENHEIMER,
H. H. Wills Physical Laboratory, University of Bristol

MS. received 22 *February* 1946

ABSTRACT. The frequencies of about 150 diatomic molecules are expressed as functions of the equilibrium distance and of the numbers of electrons in the outer shells of both atoms. The frequencies are calculated for homopolar molecules with single bonds, with multiple bonds and for molecules with strong polarity. The contributions of *s*- and *p*-electrons to the force constant are practically equal, and the same parameters are valid for all periods of the periodic system. The theoretical implications are discussed.

---

§ 1. INTRODUCTION

The properties of molecules can be derived in quantum theory from first principles. Exact solutions have been calculated for those molecules which are composed of the lightest atoms. For molecules, however, which contain atoms with many electrons, the mathematical difficulties have not yet been overcome. It is therefore of importance and of interest to find empirical rules which correlate properties like frequencies, force constants, dissociation energies and interatomic distances. The following discussion is particularly concerned with the vibration frequencies of diatomic molecules in the ground state.

Various regularities have already been suggested. Kratzer (1920) found for the halogen hydrides the relation

$$\omega^2 I = \text{const.}, \qquad \ldots\ldots(1)$$

where ω is the wave number and I the moment of inertia. Birge (1925) and Mecke (1925) proposed for other molecules a relation

$$\omega I = \text{const.} \qquad \ldots\ldots(2)$$

Morse (1929) showed that the frequencies of various molecules are related by

$$\omega r_e^3 = \text{const.}, \qquad \ldots\ldots(3)$$

where r_e is the equilibrium distance.

Badger (1934) gave a formula for the force constants

$$k(r_e - d_{ij})^3 = 1{\cdot}86 \,.\, 10^5 \text{ dyne/cm.} \qquad \ldots\ldots(4)$$

The values d_{ij} vary according to the period in the natural system of elements to which the atoms of the molecule belong, from $d_{ij} = 0{\cdot}025$ for H_2 to $d_{ij} = 1{\cdot}635$ for I_2. The force constant k is related to the frequency of the molecular vibration ω by the formula

$$k = 4\pi^2 \mu \mathbf{c}^2 \omega^2 \boldsymbol{N}^{-1} = 5{\cdot}889 \,.\, 10^{-2} \mu\omega^2 \text{ dyne/cm.}, \qquad \ldots\ldots(5)$$

where μ is the reduced mass of the molecule in units of atomic weight and $\boldsymbol{N}$ the Avogadro number.

Allen and Longair (1935) consider

$$kr_e^6 = \text{const.} \qquad \ldots\ldots(6)$$

as an adequate formula. Here again the constant varies according to the period to which both atoms belong.

Fox and Martin (1939) find that in some organic molecules

$$kr_e^5 = \text{const.} \qquad \ldots\ldots(7)$$

gives better results.

The formulae (4), (6) and (7) are certainly superior to (2) and (3), since if one compares molecules in which one atom is replaced by an isotope, the equilibrium distance is not altered and also the force constant remains unchanged, but the vibration frequency must be different owing to the different mass of the oscillator (cf. Herzfeld (1933), and Allen and Longair (1935)).

There remains, however, the difference of the forms (4) and (6), which is of theoretical importance.

§2. THE CHARGE DEPENDENCE

A further complication arises because the force constant may depend not only upon the atomic distance and upon the period, i.e. quantum number, but also upon the number of electrons or valency electrons of both atoms. Clark and co-workers (1941 *et seq.*) investigated in a series of papers whether the molecular vibration frequency is a function of the sum n of electrons of both atoms which do not belong to closed shells and came to the conclusion that

$$kr_e^6 n^{\frac{1}{2}} = \text{const.} \qquad \ldots\ldots(8)$$

and

$$kr_e^6 n = \text{const.} \qquad \ldots\ldots(9)$$

describe equally well the empirical material. Again the constants in (8) and (9) vary with the periods to which both atoms belong, e.g. in (8) from 0·107 to 9·90.

Linnett discusses another kind of charge dependence of the repulsive force in molecules, namely, the form

$$\left(\frac{z_1 + z_2}{z_1 z_2}\right)^{2/3}, \qquad \ldots\ldots(10)$$

where z_1 and z_2 are the numbers of electrons in the outer shells of the two atoms.

It is plausible to suppose that the interaction of the two atoms depends to a first approximation upon the outermost electrons only, but it seems strange that the force constant should diminish with increasing number of electrons, as is implied by formulae (8) and (9), and also by (10), if, for example, $z_1 = z_2$.

One would rather expect that the interaction, and therefore the force constant, would increase with the number of electrons involved.

If the force constant depends upon distance, and also upon number of electrons, it is obvious that one has to try to separate the two effects, either by comparing molecules with similar atomic distances or by comparing molecules whose atoms have the same number of outer electrons. The first method would give at once an indication of the form of the charge dependence. Incidentally, there are a few molecules in which the distance between the two atoms is nearly the same, e.g.

Molecule	r (A.)	k (dyne/cm.)	z_1 and z_2	
Li_2	2·6723	2·546 . 10^4	1	$\dfrac{k(I_2)}{k(Li_2)} = 6{\cdot}75$
I_2	2·667	1·718 . 10^5	7	

Lithium belongs to the first group of the periodic system, and iodine to the seventh. Therefore, one would expect that the ratio of the force constants should be a simple function of the ratio of the numbers of the outer electrons, in this case 7.

Other examples are provided by the pairs :

Molecule	r (A.)	k (dyne/cm.)	z_1	z_2	
KH	2·2244	5·555 . 10^4	1	1	$\dfrac{k(BaH)}{k(KH)} = 1{\cdot}446$
BaH	2·232	8·034 . 10^4	2	1	
CaH	2·002	9·697 . 10^4	2	1	$\dfrac{k(CaF)}{k(CaH)} = 2{\cdot}70$
CaF	2·018	2·618 . 10^5	2	7	
BeH	1·343	2·246 . 10^5	2	1	$\dfrac{k(BeF)}{k(BeH)} = 2{\cdot}56$
BeF	1·3636	5·759 . 10^5	2	7	

Assuming that the equilibrium distances of the fluorides of strontium and of barium are also nearly equal to that of their hydrides, one can compare their force constants :

Molecule	r (A.)	k (dyne/cm.)	z_1	z_2	
SrH	2·145	8·474 . 10^4	2	1	$\dfrac{k(SrF)}{k(SrH)} = 2{\cdot}72$
SrF		2·302 . 10^5	2	7	
BaH	2·232	8·168 . 10^4	2	1	$\dfrac{k(BaF)}{k(BaH)} = 2{\cdot}65$
BaF		2·166 . 10^5	2	7	

These values for the ratio of the force constants are similar to those of the previous analogous compounds of the second group, independent of the period in the natural system. This regularity is interesting in itself. But it becomes even more significant if this ratio is considered as a function of the numbers of electrons. In all cases mentioned above the force constant can be described by assuming that it depends upon the number of the outermost electrons according to the formula

$$k \sim \sqrt{(z_1 . z_2)}. \qquad \ldots\ldots(11)$$

In fact, it gives for the four cases of fluorides and hydrides of the alkaline-earth metals a theoretical ratio of $\sqrt{7} = 2{\cdot}645$ compared with an empirical average of 2·66. The ratio of the sums $z_1 + z_2$ of the outermost electrons would be 3. In the case of BaH against KH formula (11) gives a value $\sqrt{2} = 1{\cdot}4141$ compared with 1·446 empirically, and for I_2 against Li_2 $\sqrt{49} = 7$ compared with 6·75 empirically.

In all cases the experimental deviation is only a few per cent. Further evidence for this formula will be found later on.

The relation (11) is rather surprising if compared with the Coulomb law. The possibility of a wave-mechanical foundation will be discussed later.

The dependence of the force constant upon the distance can now be determined. For this purpose those molecules will first be compared which have an analogous constitution, i.e. which consist of atoms of similar valency so that the number of outer electrons in corresponding atoms remains constant. The simplest cases are provided by the molecules of hydrogen and the alkali metals. Using the method of least squares for the series H_2, Li_2, Na_2 and K_2, whose distances have been measured with spectroscopic exactitude, one gets

$$k = 2{\cdot}738 \,.\, 10^5 r_e^{-2{\cdot}46}\,\text{dyne/cm. (if } r \text{ is in A.)}, \qquad \ldots\ldots(12)$$

or, since the observed quantities are the frequency and the distance, the formula for the wave number is

$$\omega = 2{\cdot}156 \,.\, 10^3 \mu^{-\frac{1}{2}} r_e^{-1{\cdot}23}\,\text{cm.}^{-1}\ (r \text{ in A.}), \qquad \ldots\ldots(13)$$

where μ is the reduced mass in units of atomic weight.

Combined with formula (11) this gives

$$\omega = 2{\cdot}156 \,.\, 10^3 (z_1 \,.\, z_2)^{\frac{1}{4}} \mu^{-\frac{1}{2}} r_e^{-1{\cdot}23}\,\text{cm.}^{-1}\ (r \text{ in A.}) \qquad \ldots\ldots(14)$$

as a general formula. Table 1 *a* gives the values calculated according to (14) together with the distances and the observed frequencies. Those molecules are considered first whose intermolecular distance has been determined by the analysis of rotation spectra, since they represent the best measurements. Then molecules are taken whose interatomic distances have been measured by electron diffraction. The frequencies have been taken from G. Herzberg (1939), who gives a complete list of spectra measured up to 1939. Thomson and Cochrane (1939), and Maxwell (1940), give surveys of electron-diffraction measurements. In table 1 *a*, measurements of interatomic distances in polyatomic molecules have also been used on the assumption that the distance in diatomic molecules is not larger than the smallest separation of the two atoms found in polyatomic molecules with bonds of equal valency. In some cases more recent measurements by Lister and Sutton (1941) have been used. Furthermore, some molecules are inserted whose distances have been calculated by Pauling's radii (1940), which are based on X-ray crystal analysis and the additivity of covalent radii. Finally, some cases have been taken into consideration where use has been made of the fact that in the periodic system the radii of neighbouring atoms have in general similar values.

As can be seen from table 1 *a*, not only the alkali metals and the halogens satisfy formula (14), but also many other group combinations, e.g. the hydrides of elements of groups not considered above, from boron to fluorine. Particularly interesting is the long series of molecules of atoms with 4 and 7 electrons. Again the power law (13) is fulfilled. Moreover, the absolute values of the frequencies confirm the charge dependence of (11) even if these molecules are not compared with molecules with similar equilibrium distance. The transition to other groups shows also that the influence of the charge is given correctly by (11). The contributions of *s*- and *p*-electrons to the force constant are practically equal.

It is remarkable that a simple formula like (14) comprises so many cases (64 in table 1 *a*), though it contains only two constants. Moreover, since the hydrogen molecule whose properties can be calculated theoretically by wave mechanics, is also included, there is really only one constant, the power of the distance, which has to be determined.

Table 1 *a*. Molecules with single bonds : $\omega = 2\cdot156 \,.\, 10^3 (z_1 \,.\, z_2)^{\frac{1}{2}} \mu^{-\frac{1}{2}} r_e^{-1\cdot23}$ (cm.$^{-1}$).

	r_e (A.)		$\omega_{\text{obs.}}$	$\omega_{\text{calc.}}$	Error	$z_1 z_2$	
H_2	0·7414	s	4405·3	4405·3	(standard)	1×1	
Li_2	2·6723	s	361·3	344·1	2·1 %	1×1	
Na_2	3·079	s	159·2	159·4	0·1	1×1	
K_2	3·923	s	92·64	90·9	1·9	1×1	
Cl_2	1·989	s	564·9	585·3	3·6	7×7	
Br_2	2·284	s	323·2	326·6	1·0	7×7	
I_2	2·667	s	214·4	214·2	0·1	7×7	
HF	0·9166	s	4141·3	4005	3·3	1×7	
HO	0·9710	s	3728	3606	3·3	1×6	
HN	1·038	s	(3300)	3180	3·6	1×5	
HC	1·1201	s	2824	2760	2·3	1×4	
HB	1·2326	s	(2366)	2291	3·2	1×3	
BiCl	2·48	e	307·7	306	1·6	5×7	
BiBr	2·63	e	209·3	211	0·8	5×7	
CCl	1·73	e	(843·6)	845	0·2	4×7	$^2\Pi$ (CHF$_2$Cl)
SiF	1·603	s	856·7	849	0·9	4×7	$^2\Pi$
SiCl	2·00	e	535·4	536	0·1	4×7	$^2\Pi$
SiBr	2·15	e	425·4	425	0·1	4×7	$^2\Pi$
GeCl	2·08	e	408·4	403	1·3	4×7	$^2\Pi$
GeBr	2·29	e	296·6	290	2·2	4×7	$^2\Pi$
SnCl	2·32	e	352·4	353	0·2	4×7	$^2\Pi$
SnBr	2·44	e	247·7	250	0·9	4×7	$^2\Pi$
PbCl	2·43	e	303·9	304	—	4×7	$^2\Pi$
PbBr	$<$2·60	e	207·5	203	2·2	4×7	$^2\Pi$
BCl	1·73	e	830·0	811	2·3	3×7	
BBr	1·87	e	686·3	687	0·1	3×7	
AlCl	2·06	e	481·3	485	0·8	3×7	
AlBr	2·21	e	379·2	387	2·1	3×7	
AlI	2·53	e	316·1	312	1·3	3×7	
HgCl	2·23	e	293·8	285	3·0	2×7	
HgBr	2·44	e	186·2	185	0·6	2×7	
ZnI	2·42	e	223·4	216	3·3	2×7	
CdI	2·56	e	178·5	170	4·8	2×7	
HgI	2·55	e	125·0	126	0·8	(1×7)	
BiI	2·79	P	163·9	167	1·9	5×7	
GaCl	2·25		365·0	352	3·6	3×7	Wells
GaBr	2·37		263·0	263	0·1	3×7	*r*
GaI	2·54		216·4	219	1·2	3×7	Ga 1·26
InCl	2·43		316·8	299	5·6	3×7	In 1·44
InBr	2·55		221·0	213	3·6	3×7	Tl 1·47
InI	2·72		177·1	173	2·6	3×7	
TlCl	2·46		287·5	280	2·6	3×7	Cl 0·99
TlBr	2·58		192·5	191	0·8	3×7	Br 1·11
TlI	2·75		150	150	—	3×7	I 1·28
ZnF	1·59		619·5	617	0·4	2×7	*r*(ZnH)
HgF	1·74		490·8	506	3·1	2×7	HgH
BiF	1·81		510·7	513	0·5	5×7	BiH
SbF	1·82		614·0	612	0·8	5×7	TeF e

	r_e (A.)		$\omega_{\text{obs.}}$	$\omega_{\text{calc.}}$	Error	$z_1\,z_2$	
SnF	1·82		585·3	584	0·2	4×7	TeF(SnO)
PbF	2·01		507·2	508	0·2	4×7	PbO r_P
NBr	1·91		693	686	1·0	5×7	CBr e
LiK	3·30		(207)	204	1·5	1×1	r
LiRb	3·42		(185)	187	1·0	1×1	Li 1·34
LiCs	3·55		(167)	176	5·4	1×1	Na 1·54
NaK	3·50		123·3	121	1·9	1×1	K 1·96
NaRb	3·63		106·6	104	2·4	1×1	Rb 2·09
NaCs	3·75		(97)	98	1·0	1×1	Cs 2·21
Rb_2	4·18		56·78	57	0·4	1×1	$\frac{1}{2}r_e$
RbCs	4·30		49·41	49·8	0·8	1×1	Cl_2 0·99
Cs_2	4·42		41·99	42·5	1·2	1×1	Br_2 1·14
IBr	2·48		268·4	268	0·1	7×7	I_2 1·33
BrCl	2·14		(430)	456	6·0	7×7	
ICl	2·32	s	384·2	390	1·3	7×7	

s = spectroscopic
e = electron diffraction
P = Pauling radii.

At the end of table 1 *a* the radii of rubidium and caesium have been computed by using the known radii of the hydrides and by assuming that the difference of the distance in the hydrides and the radius of the alkali atom is the same as that of potassium. The calculated frequencies compare well with the observed ones, and formula (14) thus appears to be applicable from H_2 to Cs_2, i.e. from the smallest to the largest stable diatomic molecules.

§ 3. MULTIPLE BONDS

So far, only molecules with a single bond have been considered. Table 1 *b* gives the data for molecules with double bonds. The force constant varies again as $r^{-2\cdot46}$, as in molecules with single bonds described in equation (12). One new constant comes in indicating the strength of the double bond. Formula (14) is replaced by

$$\omega = 2\cdot976\,.\,10^3 (z_1\,.\,z_2)^{\frac{1}{2}} \mu^{-\frac{1}{2}} r_e^{-1\cdot23}\ \text{cm.}^{-1} \qquad \ldots\ldots(15)$$

The first factor at the right-hand side is increased by a factor 1·38. If the double bond contributed a factor 2 to the force constant, the further factor for (15) would be 1·414. The empirical factor 1·38 means that the double bond contributes a factor 1·905 to the force constant.

Table 1 *b* again gives distances, and observed and calculated wave numbers. First those molecules are listed whose interatomic distances have been determined spectroscopically. Then a series of molecules is considered whose interatomic distances are derived from Pauling's radii on the assumption of the additivity of covalent radii. It is remarkable how good the agreement is between calculated and observed wave numbers in the long series of molecules with $z_1 = 4$ and $z_2 = 6$, even where the distances have not been observed directly. Among these molecules are also those measured recently and discussed by Barrow (1944) and Jevons (1944).

The series with molecules of the nitrogen type can also be described by the same formula which has been derived from molecules which possess double bonds from the chemical point of view. The former are supposed to possess a triple bond. There is, however, no appreciable change in the relation between force constant and equilibrium distance. In this respect they belong to the same bond order.

Table 1 *b*. Molecules with multiple bonds: $\omega = 2{\cdot}976\,.\,10^3\,(z_1\,.\,z_2)^{\frac{1}{2}}\mu^{-\frac{1}{2}}r_e^{-1{\cdot}23}$ (cm.$^{-1}$)

	r_e (A.)		$\omega_{\text{obs.}}$	$\omega_{\text{calc.}}$	Error	$z_1 z_2$	
CO	1·1284	s	2168	2168	(standard)	4×6	
SiO	1·510	s	1242	1243	0·1 %	4×6	
TiO	1·620	s	1008	1050	4·2	4×6	$^3\Pi_u$
GeO	1·651	s	985·6	980	0·6	4×6	
SnO	1·838	s	822·4	829·9	0·9	4×6	
CS	1·563	s	1285·1	1301	1·2	4×6	
PbS	2·395	s	428·1	422·5	1·3	4×6	
N_2	1·095	s	2359·6	2251	4·6	5×5	
P_2	1·898	s	780·4	773·1	0·9	5×5	
NP	1·491	s	1337	1312	1·9	5×5	
CN	1·172	s	2068·7	2036	1·6	4×5	
CP	1·5622	s	1240	1237	0·2	4×5	
SiN	1·572	s	1152	1181	2·5	4×5	
BO	1·205	s	1885·4	1907	1·2	3×6	
NO	1·150	s	1907	1940	1·7	4×5	$^2\Pi$
PO	1·447	s	1231	1230	0·1	4×5	$^2\Pi$
NS	1·50		1220	1226	0·5	4×5	$^2\Pi$ r(SO)
AsO	1·61		960·6	964	0·3	4×5	$^2\Pi$ (SeO)
CSe	1·735	P	1036	1036	—	4×6	
SiS	2·01	P	749·5	726	3·1	4×6	r_{Pauling}
SiSe	2·14	P	580·0	569	2·0	4×6	C 0·665
SiTe	2·34	P	480·4	483	0·5	4×6	Si 1·07
GeS	2·06	P	575·8	576	—	4×6	Ge 1·12
GeSe	2·19	P	406·8	406	0·2	4×6	Sn 1·30
GeTe	2·39	P	323·4	329	1·7	4×6	Pb 1·46
SnSe	2·37	P	332·0	329	0·9	4×6	
SnTe	2·57	P	259·5	262	1·0	4×6	O 0·55
SnS	2·24	P	487·7	487	0·1	4×6	S 0·94
PbO	2·01	P	721·8	724	0·3	4×6	Se 1·07
PbSe	2·53	P	277·4	277	0·1	4×6	Te 1·27
PbTe	2·73	P	211·8	216	2·0	4×6	
							r
As_2	2·19		429·4	414	3·5	5×5	(GeSe)
Sb_2	2·57		272·8	270	1·0	5×5	(SnTe)
Bi_2	2·92	P	173·7	174	0·2	5×5	
AsN	1·65		1068	1049	1·8	5×5	(GeO)
SbBi	2·73		220	221	0·6	5×5	(PbTe)

In this series the frequencies of some molecules are computed according to (15) on the assumption that the distance of, for example, Sb_2 is the same as in SeTe, the sum of the radii of the preceding and of the following elements.

All these molecules with multiple bonds obey the same power law, $k \sim r^{-2{\cdot}46}$, as molecules with single bonds.

§ 4. POLAR MOLECULES

There is another class of molecules ; they do not obey the $r^{-2\cdot46}$ relation. Most hydrides, and also the halides of the alkali and of the alkaline-earth metals, belong to this class. It is easy to find a simple formula for these molecules also, again with only two constants, of the same form as (14), i.e. of the type $k = Cr^{s}$.

Formula (14) has to be replaced by

$$\omega = 2{\cdot}065 \,.\, 10^{3} (z_1 \,.\, z_2)_{\frac{1}{2}} \mu^{-\frac{1}{2}} r_e^{-0\cdot92} \text{ cm.}^{-1} \ (r \text{ in A.}). \qquad \ldots\ldots (16)$$

The values for the two constants have been derived by using NaH and CsH as standard molecules. The vibrations of KH and RbH are represented by the same formula. Moreover, not only BaH but also the other hydrides of the alkaline-earth metals obey formula (16) and confirm the charge dependence of formula (11) within the experimental error. It is worth noticing that the vibrations of alkali halides can also be described by the same formula, taking $z_1 = 1$; $z_2 = 7$.

One important conclusion can be drawn from this fact : according to (16), the wave number is a function of the product of the numbers of outer electrons. From the agreement of calculated and observed wave numbers one can conclude that the molecules are built up not by ions but by neutral atoms. If this is true even for the alkali halides, it is also true for the other molecules.

Most of the molecules whose frequencies obey formula (16) are known to have a strong dipole moment. This can be considered as responsible for the fact that the force constant of these molecules varies much more slowly with distance than that of molecules whose dipole moment is small or zero.

The group of molecules of the oxygen type belongs also to this series. They are known to have a $^{3}\Sigma$ ground state. The parallelism of the spins seems to be sufficient to create a contribution to the force constant.

In table 2 all known hydrides of metals have been listed for the sake of completeness. As in table 1, the numbers of outer electrons used for the calculations of the frequencies correspond to the maximum valencies of the atoms. However, the frequencies of the hydrides of Cu, Ag and Au, which belong to the transition elements, are much higher than expected from equation (16). This is probably due to the effect of the last attached *d*-electrons at small distances. One could use equation (16) to determine an effective number of electrons which allows the best description of the empirical frequencies in these three cases. They are shown in brackets in table 2. The halides of copper behave normally.

Table 2. $\omega = 2{\cdot}065 \,.\, 10^{3} (z_1 \,.\, z_2)^{\frac{1}{2}} \mu^{-\frac{1}{2}} r_e^{-0\cdot92}$ (cm.$^{-1}$)

	r (A.)	$\omega_{\text{obs.}}$	$\omega_{\text{calc.}}$	Error	$z_1\ z_2$
HNa	1·8875 s	1170·8	1170·8	(standard)	
HCs	2·494 s	890·7	890·7	(standard)	
HK	2·244 s	983·3	989·8	0·7 %	1×1
HRb	2·368 s	936	935·8	—	1×1
HLi	1·5956 s	1406	1430	1·7	1×1
HBe	1·343 s	2058	1965	4·5	1×2
HMg	1·7302 s	1495	1506	0·7	1×2
HCa	2·002 s	1299	1307	0·6	1×2
HSr	2·1457 s	1207	1218	0·9	1×2

Table 2—*continued*

	r (A.)	$\omega_{\text{obs.}}$	$\omega_{\text{calc.}}$	Error	$z_1 z_2$	
HBa	2·232 s	1172	1173	0·1	1×2	
HZn	1·5947 s	1608	1604	0·2	1×2	
HCd	1·762 s	1431	1458	1·9	1×2	
HHg	1·741 s	1387	1472	6·1	1×2	
HTl	1·870 s	1391	1525	9·6	1×3	
HPb	1·839 s	1565	1664	6·3	1×4	
HBi	1·809 s	1699	1786	5·1	1×5	
HAl	1·6461 s	1683	1744	3·6	1×3	
HSi	1·521 s	(2080)	2109	1·4	1×4	
HP	1·433 s	(2380)	2244	5·7	1×5	
HCo	(1·543) s	(1890)	1831	3·1	1×3	
HNi	1·475 s	(1927)	1904	1·2	1×3	
HCu	1·463 s	1940	1921	1·0	(1×3)	
HAg	1·618 s	1760	1785	1·4	(1×3)	
HAu	1·5239 s	2306	2275	1·3	(1×7)	
HCl	1·2747 s	2980	2713	9·0	1×7	
HBr	1·414 s	2650	2444	7·8	1×7	
HI	1·604 s	2310	2174	5·9	1×7	
O_2	1·2076 s	1580	1497	5·3	6×6	$^3\Sigma$
S_2	1·89 e	725·8	701	3·4	6×6	$^3\Sigma$
Se_2	2·16 e	392·5	392	0·1	6×6	$^3\Sigma$
Te_2	2·59 e	251·5	262	4·2	6×6	$^3\Sigma$
SO	1·4935 s	1123·7	1066	5·1	6×6	$^3\Sigma$
SeO	1·61 e	908·9	891	2·0	6×6	$^3\Sigma$
TeO	1·82	796·1	768	3·5	6×6	
NaCl	2·51 e	380	385	1·3	1×7	
NaBr	2·64 e	315	325	3·2	1×7	
NaI	2·90 e	286	285	0·5	1×7	
KCl	2·79 e	280	303	8·2	1×7	
KBr	2·94 e	231	242	4·8	1×7	
KI	3·23 e	212	208	1·9	1×7	
BeF	1·362 s	1266	1211	4·3	2×7	
CaF	2·018 s	587·3	579·4	1·3	2×7	r of
SrF	2·146	500·1	498	0·4	2×7	SrH
BaF	2·23	468·9	465	0·8	2×7	BaH
MgCl	<2·51	466·0	452	3·0	2×7	NaCl
MgBr	~2·64	373·2	379	1·6	2×7	NaBr
CaCl	<2·79	364·5	358	1·8	2×7	KCl
SrCl	~2·79	301·1	309	2·6	2×7	KCl
BaCl	~2·79	279·2	292	4·6	2×7	KCl
CaBr	~2·94	(280·2)	292	4·2	2×7	KBr
SrBr	~2·94	(212)	232	(9·3)	2×7	KBr
BaBr	~2·94	(192)	204	(6·1)	2×7	KBr
CuF	1·759 s	619·5	619·2	0·1	2×7	
CuCl	2·13 e	417·0	418	0·2	2×7	
CuBr	2·25 e	314·1	319	1·6	2×7	
CuI	2·40 e	264·8	268	1·2	2×7	
C_2	1·3121 s	1641·7	1600	2·5	3×3	$^3\Pi_u$

At the end of table 2, the vibrations of a few molecules have been considered whose intermolecular distances have not been measured directly but have been inferred indirectly from molecules of similar size. The agreement between calculated and empirical wave numbers is sufficient to show that these molecules also belong to the series of polar molecules which obey formula (16).

There remains the possibility that there are also molecules with a double bond which belong to this series with the smaller power law. The molecule C_2 appears to be of this type. This molecule is known to have as ground state a $^3\Pi$ configuration. In molecules with triplet states the eigenfunction must be antisymmetric and, therefore, have a node at the centre of the line which goes through both nuclei. This means that in C_2 one pair of charges plays practically no rôle in a region where normally the interaction of the two atoms is greatest. Thus it is understandable that only three electrons of each atom are effective, and on this assumption (allowing for a factor 2 for the force constant of the double bond), the frequency of C_2 is given satisfactorily by (16).

A similar phenomenon seems to occur in the four molecules NO, PO, AsO and NS of table 1 *b*. All of them have doublet ground states ($^2\Pi$). This can mean that all electrons apart from the odd one form together either a singlet or a triplet state. The latter possibility would again give rise to a practical non-participation of a pair of electrons in the critical region between the two atoms. Empirically these four molecules show among themselves the correct dependence $r_e^{-1 \cdot 23}$ upon the distance, but the charge product seems to be diminished by just one unit in each atom. This phenomenon seems to occur only with Π states with their pronounced directional property.

§ 5. DISCUSSION

In the foregoing it has been shown that for a substantial part of all molecules whose frequencies in the ground state have been measured, the force constant can be described by a simple function of the equilibrium distance. The formulae depend only on two constants. The difference between the first constant for the frequencies in formulae (14) and (16) is only 5 %. The $r^{-1 \cdot 23}$ power relation is fulfilled by about one hundred molecules ; the power rule $r^{-0 \cdot 92}$ by about fifty.

The $r^{-1 \cdot 23}$ relation covers the range from the highest observed wave number, 4405 cm.$^{-1}$ (H_2), to the smallest wave number, 42 cm.$^{-1}$ (Cs_2). In the series of molecules with multiple bonds the same power law extends from the molecule BO to $^{209}_{83}Bi_2$, the heaviest stable diatomic molecule.

The difference between calculated and observed frequencies is for the one hundred molecules of table 1, on the average 1·6 %, and often less than the uncertainty due to the errors in the measurements of the equilibrium distance.

There is practically no difference between the effect of *s*- and *p*-electrons, nor is the period of the elements of any importance.

The dependence upon the distance is smaller than would be predicted by formulae (4), (6) or (7).

The fact that the variations of long series of molecules can be described by formulae with only two constants is not only remarkable in itself but must indicate a simple but fundamental physical fact.

It is understandable that the force constant between two atoms depends, to a first approximation, only on the number of electrons which are in the outermost shells of both atoms. Ample evidence for, and confirmation of, formula (11) have been found in tables 1 and 2. The form of (11), the proportionality of the force constant to the geometric mean number of electrons, is, however, new and surprising.

In the following it will be shown how this relation can be understood within the framework of quantum mechanics. In wave mechanics the density of the charge, or the probability of finding one electron, is given by

$$\rho = \psi\psi^* d\tau \qquad \ldots\ldots(17)$$

and the eigenfunction is normalized by

$$\int \psi\psi^* d\tau = 1. \qquad \ldots\ldots(18)$$

If there were z_1 electrons with the same eigenfunction, the probability of finding one electron at a certain place would be z_1 times as great and, therefore,

$$\int \psi\psi^* d\tau = z_1. \qquad \ldots\ldots(19)$$

The Pauli principle does not allow more than one electron to be in a given state. The electrons, however, which belong to the same shell, i.e. have the same principal quantum number, have similar radial eigenfunctions, at least the *s*- and *p*-electrons. In particular the mean distance from the nucleus is rather similar. The probability of finding one electron in the outermost shell will therefore be z_1 times as great as if there were only one electron there. In fact, the method of the self-consistent field of Hartree is based on this condition.

Generally, the function which represents the solution for the whole system of electrons can be considered to be the result of the superposition of the eigenfunctions of all the electrons which belong to the system

$$\Psi_1 = \Sigma c_i \psi_i, \qquad c_i^2 = 1. \qquad \ldots\ldots(20)$$

As a matter of fact, the electrons are indistinguishable. In a system of electrons, each electron can belong to each state which is occupied. The probability of finding one electron at a certain position is equal to the sum of the probabilities due to the eigenfunctions which belong to the occupied states

$$\Psi_1 \Psi_1^* d\tau = \Sigma \psi_i \psi_i^* d\tau. \qquad \ldots\ldots(21)$$

If the eigenfunctions are orthogonal, condition (19),

$$\int \Psi_1 \Psi_1^* d\tau = \int \Sigma \psi_i \psi_i^* d\tau = z_1, \qquad \ldots\ldots(22)$$

is fulfilled, and the orthogonalization is always possible. The condition of the form (22) corresponds to the quantization of the amplitudes introduced by Dirac (1927) in the theory of radiation.

If there are two atoms, equation (22) has to be fulfilled also for the second atom,

$$\int \Psi_2 \Psi_2^* d\tau = z_2. \qquad \ldots\ldots(23)$$

If two separated atoms approach one another, the two outer shells will begin to overlap. This gives rise to a perturbation. The energy of the whole system becomes a function of the distance of the two nuclei. The force constant is equal to the second derivative of the energy with respect to the interatomic distance, taken at the equilibrium distance.

According to equation (11) the force constant is empirically proportional to the geometric mean of the numbers of the outermost electrons,

$$k \sim \sqrt{(z_1 . z_2)}. \qquad \ldots\ldots(24)$$

With regard to equations (22) and (23), this fact obtains a simple theoretical meaning :—The force constant is proportional to the mixed density of the electrons

$$k \sim |\Psi_1 \Psi_2|. \qquad \ldots\ldots(25)$$

In the one-electron problem, H_2^+, the binding energy, is essentially due to the overlap integral $\int \frac{1}{r_1} \psi_1 \psi_2 \, d\tau$, if use is made of the method of linear combination of atomic orbitals, and also the force constant is, at least to a first approximation, theoretically proportional to the mixed density. The result of (25) appears as a natural extension of this fact.

The repulsive force contributes more to the magnitude of the force constant, since it varies more quickly with the distance than the force of attraction. At the equilibrium distance the repulsive force and the attractive force are equal in magnitude. The force constant is due to the difference of the derivatives of these forces. One cannot derive the law of one of these forces alone from the rules connecting force constant and distance given by equations (14) to (16), but the variation of the force constant with distance gives at least a lower limit for the variation of the repulsive force, and conversely, a given law for the repulsive force should give an upper limit for the variations of the force constant.

The most important contribution to the repulsive force is due to the pressure of the electronic gas which is built up by the electrons of both atoms, particularly of those which participate in the binding. To get an indication of the magnitude of these forces one has to consider the electronic gas confined to the volume of the molecule. The energy of the electronic gas is then, according to Fermi,

$$E = \frac{3}{40} \frac{h^2}{m} \left(\frac{3}{\pi}\right)^{2/3} \frac{z^{5/3}}{V^{2/3}}. \qquad \ldots\ldots(26)$$

The contribution to the force constant is determined by the second derivative with respect to the atomic distance. The order of magnitude is given by $\hbar^2/mr^4$. This yields for $r = 1$ A. an order of magnitude of 10^5 dyne/cm. This is necessary and sufficient to satisfy the requirements of equation (12), which describes the empirical facts.

The occurrence of the mass of the electrons in the expression $\hbar^2/mr^4$ shows that the electrons are responsible for the magnitude of the force constant, though the frequencies which are measured are due to the vibrations of the total masses of the atoms of the molecule.

The variation of the atomic distance during vibration leads to a variation of the volume of the electronic gas proportional to the atomic distance. As a result of equation (26) the variation of the energy is then proportional only to $r^{-2/3}$, where r is the distance between the two atoms, and taking the second derivative leads to $r^{-2 \cdot 67}$ for the variation of the force constant. Owing to the effect of binding forces and of the non-existence of rigid limits for the electronic gas, the actual variation of the force constant will be slower. Comparison with the empirical formula $r^{-2\cdot 46}$ of equation (12) shows that this consideration leads at least to a good approximation and yields the right order of magnitude for the exponent of the distance r in equation (12).

The relations for the force constant of diatomic molecules which have been derived in this paper can be applied also to polyatomic molecules, and can be compared with other properties of matter, e.g. the compressibilities and thermal expansion of solid bodies. This will be done in later papers.

REFERENCES

Allen and Longair, 1935. *Phil. Mag.*, **19**, 1032.
Badger, 1934. *J. Chem. Phys.*, **2**, 128.
Barrow, 1944. *Proc. Phys. Soc.*, **56**, 204.
Birge, 1925. *Phys. Rev.*, **25**, 240.
Clark and Webb, 1941. *Trans. Faraday Soc.*, **37**, 293.
Dirac, 1927. *Proc. Roy. Soc.*, **114**, 243.
Fox and Martin, 1939. *J. Chem. Soc.*, 884.
Herzberg, 1939. *Molecular Spectra and Molecular Structure.*
Herzfeld, 1933. *Handb. d. Phys.*, xxiv/2, 98.
Jevons, 1944. *Proc. Phys. Soc.*, **56**, 211.
Kratzer, 1920. *Z. Phys.*, **3**, 289.
Linnett, 1942. *Trans. Faraday Soc.*, **38**, 1.
Lister and Sutton, 1941. *Trans. Faraday Soc.*, **37**, 393, 406.
Maxwell, 1940. *J. Opt. Soc. Amer.*, **30**, 375.
Mecke, 1925. *Z. Phys.*, **32**, 823.
Morse, 1929. *Phys. Rev.*, **34**, 57.
Pauling, 1940. *Nature of the Chemical Bond.*
Skinner and Sutton, 1944. *Trans. Faraday Soc.*, **40**, 164.
Thomson and Cochrane, 1939. *Theory and Practice of Electron Diffraction.*
Wells, 1944. *Structural Inorganic Chemistry.*

TIDAL EFFECTS ON THE PRODUCTION OF MESONS IN THE ATMOSPHERE

By A. W. MAILVAGANAM,
Colombo, Ceylon

Communicated by Prof. P. M. S. Blackett, F.R.S.; *MS. received* 4 *March* 1946

ABSTRACT. The theory of forced oscillations, magnified by resonance, of the earth's atmosphere with a period of 12 solar hours, as worked out by Pekeris, is applied to calculate the amplitude of oscillation of the layer in the atmosphere in which mesons are produced by primary incident particles. The calculations show that, on account of these oscillations and of the finite life-time of mesons, a semi-diurnal variation in meson intensity with an amplitude of the order of 0·14 % may be expected.

The main features of the semi-diurnal barometric oscillation attributed to tides in the atmosphere are now well known. It has two components, a solar and a lunar, the amplitude of the former being about 15 times that of the latter. To explain the solar component, Kelvin made the suggestion that the atmosphere has a free period of nearly 12 hours, so that the solar semi-diurnal tide is amplified by resonance. The free and forced oscillations of the atmosphere have been studied by G. I. Taylor (1936) and by C. L. Pekeris (1937, 1939). Taylor has shown from a study of the propagation of waves of explosion that the atmosphere has a free period of oscillation of about 10·5 hours. Pekeris, using the known distribution of temperature within the first 60 kilometres of the atmosphere, and by assuming the simplest type of temperature distribution above 60 kilometres consistent with the existence of a free semi-diurnal oscillation in the atmosphere, has shown that the resulting free oscillation possesses the characteristics required by the "dynamo" theory of the diurnal variation of the earth's magnetic field. It also followed from Pekeris' work, in agreement with Taylor's results, that the atmosphere so "adjusted" possesses yet another period of free oscillation of a period of about 10·5 hours.

In an appendix to his second paper, Pekeris has given curves showing the variation of the semi-amplitudes of the vertical displacements of a particle with height above the ground for the two modes of oscillation. We note from the curve for the 12-hour period that these amplitudes become appreciable at altitudes above 10 kilometres. It may be recalled in this connection that Appleton and Weekes (1939), using radio methods of upper-atmospheric exploration, have detected a lunar tide in the Kennelly-Heaviside layer. They found in the experimentally determined expression for the equivalent height of the lower boundary of this layer a simple harmonical term, $0{\cdot}93 \sin(2t' + 112)$ km. They concluded that the tide is semi-diurnal, in phase with the lunar barometric oscillations as observed by Chapman at Greenwich, and of the order of 1 kilometre.

Since, according to Pekeris, the amplitude of the oscillations of a particle increases rapidly with altitude, it is of interest to examine the extent to which the layer in which mesons are produced in the atmosphere oscillates owing to the semi-diurnal solar tide. For it is known from the work of Euler and Heisenberg (1938) that mesons are radioactively unstable, with a mean time of decay of about 2×10^{-6} sec. in a system of co-ordinates in which they are at rest. In consequence, if the layer of production rises in the atmosphere, then fewer mesons would reach the earth, owing to the increased path the mesons have now to travel. Blackett (1938) has successfully used these considerations to explain (i) the temperature effect of cosmic rays, and (ii) the observations of Auger, Ehrenfest, Freon and Fournier (1937), who found that the absorption of inclined cosmic rays was greater than that of vertical rays under the same thickness of absorber. Blackett showed that both effects could be explained in terms of meson decay, and deduced from the available data that the mean free path (decay) of the mesons in the atmosphere is about 24 km.

In this work account will be taken only of the solar semi-diurnal tide, the corresponding lunar tide being much smaller in comparison. The work of Pekeris (1937) will be followed closely and the notation employed by him used. In order to simplify the calculations the following stratification of the atmosphere, slightly different from that employed by Pekeris, is assumed. A rough check indicates that no appreciable error is introduced by this simplification.

The numbers 0, 1, 2, 3, 4 denote the boundaries between the sections of the atmosphere, 0 referring to the ground level. From the work of Hamilton, Heitler and Peng (1943) we take the meson-producing layer as lying between equivalent water-thicknesses of 15 and 30 cm. below the top of the atmosphere.

In centimetres of mercury these will correspond to 1·1 to 2·2. The summer heights in the atmosphere corresponding to these pressures are, according to Humphreys (1929), 29 and 24·5 km. respectively. In order to estimate the order of magnitude of the rise of the layer of meson production it would be sufficient to use either the summer or the winter values. It will be seen that this layer lies in the stratosphere (i.e. in region 1–2).

	Layer	Position in atmosphere (km.)	Temp. (° K.)	Remarks
Troposphere	0–1	0–10	287–230	Negative lapse rate
Stratosphere	1–2	10–37·5	230	Constant temp.
Layer of abnormal reflection of sound waves	2–3	37·5–60	230–270	Positive lapse rate
Adjusted top of atmosphere	3–4	60–80	270–215	Negative lapse rate
	4	>80	215	Constant temp.

For our purpose it is necessary to calculate the amplitude of the semi-diurnal tidal oscillation of the isobars passing through the heights under reference. The oscillations of any particular level have been calculated by Pekeris and represented graphically (figure 3, curve Z_2, of his paper (1939)). The magnitudes of the two kinds of oscillation must necessarily be different, the former being smaller owing to the heaping-up of the atmosphere under tidal influence. Using the notation employed by Pekeris, we have, if p is the pressure variation and h the displacement of an isobar under tidal influence,

$$\frac{\partial p}{\partial t}=\sigma p, \quad p=-g\rho h,$$

where $\sigma=2\omega$; ω is the angular velocity of the earth's rotation, ρ the density of air at the level concerned, and g the acceleration due to gravity. These two equations, along with equation 34 (Pekeris, 1937),

$$\frac{1}{\rho}\frac{\partial p}{\partial t}=-\frac{\gamma}{2}gH\{C(1+v)e^{\lambda_1 y}+D(1-v)e^{\lambda_2 y}\},$$

yield

$$h=\frac{\gamma}{2\sigma}H\{C(1+v)e^{\lambda_1 y}+D(1-v)e^{\lambda_2 y}\},$$

where y is given by the relation $\left(\frac{287}{5\cdot7}-Z\right)\times10^5$, Z being the height of a layer measured in kilometres, and γ is the ratio of the two specific heats of air. In the derivation of these equations we have omitted writing the imaginary quantity i, which only takes into account the phase of the oscillation. Furthermore, we have assumed ρ, the density of the air, to be constant over the distance h and equal to the value on the isobar concerned. The constants appearing in this equation have been calculated by the method indicated by Pekeris (1939, appendix), and are as follows:—

$$C=\ 0\cdot0794\times10^{-5}, \quad D=-0\cdot975\times10^{-5}$$
$$\lambda_1=-0\cdot0634\times10^{-5}, \quad \lambda_2=-0\cdot0851\times10^{-5}$$
$$V=\ 0\cdot1483, \quad H=\ 7\cdot87\times10^5\,\text{cm.}$$

Using these values, we find that the limits of the production layer estimated by Hamilton, Heitler and Peng are raised by $4{\cdot}12\times10^3$ cm. and $2{\cdot}76\times10^3$ cm. respectively at tidal maximum. Comparison with the curve given by Pekeris shows that these displacements of the isobars are smaller than, but of the same order of magnitude as, the displacements of layers at the respective heights.

From the above calculations we find that the lower surface of the production layer is raised at tidal maximum by $3{\cdot}44\times10^3$ cm. (mean), the mean height of the surface being 27 km. Using Blackett's estimate of the mean free path (decay) of the meson, we obtain the following equations for the mesons reaching the earth at zero tide and maximum tide respectively:—

$$N = N_0 e^{-2{\cdot}7/2{\cdot}4}, \quad N' = N_0 e^{-2{\cdot}7034/2{\cdot}4},$$

where N_0 is the number of mesons produced in the layer. We thus get as a lower limit for the percentage variation in meson intensity at tidal maximum

$$\frac{N-N'}{N}\times 100 = 0{\cdot}14\%.$$

According to Simpson (1918), the semi-diurnal pressure maxima due to the solar tide occur at 10 a.m. and 10 p.m. local time. The minima in the meson intensity due to solar tides should therefore occur at these times. The tidal theory developed by Pekeris may be considered satisfactory in many respects; but it fails to explain why the semi-diurnal maxima occur at 10 a.m. and 10 p.m. (local time).* It may also be objected that the assumed temperature variation for the top of the atmosphere, which is by no means unique, does not fit in with the data obtained by the radio-exploration of the upper atmosphere.

In conclusion I would take this opportunity to thank Mr. M. Kidnapillai, of the Ceylon Technical College, for much valuable assistance in the calculations, and Mr. V. Appapillai for helpful discussions.

REFERENCES

Appleton, E. V. and Weekes, K., 1939. *Proc. Roy. Soc.*, A, **171**, 171.

Auger, P., Ehrenfest, P., Freon, A. and Fournier, A., 1937. *C.R. Acad. Sci., Paris*, **204**, 257.

Blackett, P. M. S., 1938. *Phys. Rev.*, **54**, 973.

Blackett, P. M. S., 1938, *Nature, Lond.*, **142**, 992.

Chapman, S., 1924. *J. Roy. Met. Soc.*, **50**, 165.

Euler, H. and Heisenberg, W., 1938. *Ergeb. d. Exakt. Nat.*, **17**, 1.

Hamilton, J., Heitler, W. and Peng, H. W., 1943. *Phys. Rev.*, **64**, 78.

Humphreys, W. J., 1929. *Physics of the Air* (London), Table III.

Pekeris, C. L., 1937. *Proc. Roy. Soc.*, A, **158**, 650.

Pekeris, C. L., 1939. *Proc. Roy. Soc.*, A, **171**, 434.

Simpson, G. C., 1918. *Quart. J. Roy. Met. Soc.*, **44**, 1.

Taylor, G. I., 1936. *Proc. Roy. Soc.*, A, **156**, 318.

Rossi, Heberry and Hoag, 1940. *Phys. Rev.*, **57**, 461.

---

* This has, however, been explained by Chapman (1924).

NOTE ON MEASUREMENTS OF GLASS ABSORPTION

By T. SMITH, F.R.S.,
National Physical Laboratory

MS. received 27 *April* 1946

ABSTRACT. The theory of four methods of measuring the transmissivity of optical glass is given, with special reference to the effect of surface reflexion. The methods include one recently described by A. Maréchal, and two which have been used for industrial measurements. The accuracy attained by all four methods is similar provided the appropriate conditions, which are more stringent in some cases than others, are satisfied.

One of the chief requirements of glass from which optical instruments are to be made is a high transmission factor. Great attention is given to the purification of the raw materials, particularly important being a very low iron content in the sand. Correspondingly important is the accuracy of the measurements from which the transmissivity of the glass is derived.

The loss of light in passing through glass is attributable to two distinct causes. Within the glass, light is absorbed, and at the surfaces some is reflected. The surface losses depend on the state of the surface; a freshly polished surface may differ from a clean surface that has been exposed to the air for some days, and a bloomed surface will, in general, differ appreciably from an untreated surface. For these and other reasons it is desired to separate the two effects, and particularly to determine the true loss due to absorption within the glass, from which the transmissivity, i.e. the proportion of light transmitted through a thickness of 1 cm. of glass, is obtained. To obtain accurate values the specimens examined should give a long glass path, so that any errors in the measurements will have a very small effect on the value found for the transmissivity. For example, if the transmittance of a specimen 20 cm. long is 80 ± 1 per cent, the transmissivity is $0{\cdot}9889 \pm 0{\cdot}0006$, so that the accuracy is of the order of one part in two thousand, though the direct measurements are uncertain to about one per cent.

Before considering methods of measurement it is convenient to establish the relations between the glass and surface properties and the measurable quantities. At each surface some of the incident light is reflected and some transmitted, but it is assumed that none is absorbed; let the ratio of the reflected to the transmitted light be ϵ. Within the glass the fraction τ is transmitted and none is reflected.

The measurable quantities are the intensities of the beams of light entering and leaving the specimen. It is convenient to consider light travelling in both directions on each side of the specimen. Let the light intensities of the entering and leaving beams on one side be denoted by i and j respectively, and the beams on the other side travelling in the same directions be denoted by i' and j' respectively. The emergent light i' is composed of the transmitted part of i and the reflected part of j', and similarly j is composed of the reflected part of i and the transmitted part of j'. Let t and r denote the transmission and reflexion factors of the complete assembly. It is known that t is independent of the direction in which the light travels, and that the values of r for both directions are equal if the system is symmetrical or when there is no absorption. The relations between the beams may be written concisely

$$(i'\,j')\begin{pmatrix}1 & \cdot\\ -r & t\end{pmatrix} = (i\,j)\begin{pmatrix} t & -r\\ \cdot & 1\end{pmatrix},$$

so that

$$(i'\,j')=(i\,j)\begin{bmatrix} t-\dfrac{r^2}{t} & -\dfrac{r}{t}\\[2ex] \dfrac{r}{t} & \dfrac{1}{t}\end{bmatrix}. \qquad \ldots\ldots(1)$$

When light falls on a succession of media the last equation shows that the resultant effect is obtained by multiplying in their proper order matrices of the form shown on the extreme right of this equation. If we note that when there is no absorption $t+r=1$, and that this matrix then takes the form

$$\begin{pmatrix}1-\epsilon & -\epsilon\\ \epsilon & 1+\epsilon\end{pmatrix}$$

we see that for the glass specimen under consideration

$$\begin{bmatrix} t-\dfrac{r^2}{t} & -\dfrac{r}{t}\\[2ex] \dfrac{r}{t} & \dfrac{1}{t}\end{bmatrix} = \begin{bmatrix}1-\epsilon & -\epsilon\\[2ex] \epsilon & 1+\epsilon\end{bmatrix}\begin{bmatrix}\tau & \cdot\\[2ex] \cdot & \dfrac{1}{\tau}\end{bmatrix}\begin{bmatrix}1-\epsilon & -\epsilon\\[2ex] \epsilon & 1+\epsilon\end{bmatrix}. \qquad \ldots\ldots(2)$$

Clearly to obtain the transmittance τ two quantities must be measured.

When the refractive index μ of the glass is known, and the surfaces are freshly polished, ϵ has the value $\dfrac{(\mu-1)^2}{4\mu}$. From the value of $\dfrac{1}{t}$ in equation (2) we obtain the quadratic equation

$$\tau = t\{(1+\epsilon)^2-(\tau\epsilon)^2\} \qquad \ldots\ldots(3)$$

for τ. The percentage accuracy in τ is substantially equal to that in t. The experimental procedure is to determine the ratio of i' to i when $j'=0$. In practice it is sufficient to know the type of glass being measured, as this will determine ϵ with ample accuracy provided a freshly polished specimen is used. If the surfaces are well bloomed ϵ will be small enough to be neglected.

A method that has been extensively employed is to make transmission measurements on two specimens of the same glass, one very thick and the other thin. Assuming that their surfaces are in a similar state, equation (3) gives, with ample accuracy,

$$\frac{\tau_1}{\tau_2}=\frac{t_1}{t_2}\cdot\frac{1-\{t_1\epsilon(1+\epsilon)\}^2}{1-\{t_2\epsilon(1+\epsilon)\}^2},$$

where a very rough value of ϵ is adequate. The left side of this equation represents the transmission factor for a thickness of glass equal to the difference of the thicknesses of the two specimens. As the percentage errors in the values of t are algebraically subtractive, the corresponding error in τ may be twice that made in a measurement of t.

A third method that is less familiar is to determine both r and t experimentally. ϵ will disappear from equation (2) if we pre-multiply throughout by $(1\ \ 1)$ and post-multiply by $\begin{pmatrix}-1\\1\end{pmatrix}$. This gives

$$\frac{(1-r)^2}{t}-t=\frac{1}{\tau}-\tau,$$

a quadratic equation for τ. When t and r are known, τ can therefore be found whatever the condition of the surfaces may be so long as there is no surface absorption. The two surfaces may be dissimilar or non-uniform in their properties, though in the latter event the same areas must receive the light in all measurements. If the errors in the measurements of t and r are δt and δr, the corresponding error in τ is $\delta\tau$ where,

$$\left(\frac{1}{\tau^2}+1\right)\delta\tau=\left\{\left(\frac{1-r}{t}\right)^2+1\right\}\delta t+2\,\frac{1-r}{t}\,\delta r,$$

or the error in τ is approximately the sum of those in t and r. As r tends to be small, the error in its value will normally be of little importance, and the error in τ is substantially equal to that in t.

A new method has been proposed recently by Maréchal (1944). He measures the brightness of the reflected beam when the glass specimen is backed by a mirror having a reflexion factor ρ. If R is the observed reflexion factor of the combined system, equation (1) gives

$$(1\quad R)\begin{bmatrix} t-\dfrac{r^2}{t} & -\dfrac{r}{t} \\[2ex] \dfrac{r}{t} & \dfrac{1}{t}\end{bmatrix}\begin{bmatrix}-\rho\\[2ex] 1\end{bmatrix}=0$$

or

$$\frac{R-r}{\rho}=\frac{t^2}{1-r\rho},$$

since $j'=\rho i'$. When ϵ, and therefore also r, is small enough to be neglected the equation reduces to

$$\tau^2=\frac{R}{\rho}.$$

This equation will be applicable when the surfaces of the specimen are carefully bloomed. The relative accuracy of the value of τ is equal to that of a single reflexion factor.

When ϵ is not negligible, substitution for r and t from (2) leads to a quadratic equation in τ^2 which may be written

$$\tau^2\rho=$$
$$[(1+\epsilon)\{1+\epsilon(1-\rho)\}-\tau^2\epsilon\{\rho+\epsilon(1-\rho)\}]\,[(1+\epsilon)\{R-\epsilon(1-R)\}-\tau^2\epsilon\{1-\epsilon(1-R)\}],$$

so that τ^2 is a symmetrical function of R and $1/\rho$. In both factors the coefficients of ϵ tend to be small and those of ϵ^2 decidedly smaller, so that the correction required owing to surface reflexion is substantially represented by first-order terms. This gives

$$\tau^2 = \frac{R}{\rho}\left[1-\epsilon\left\{\frac{(1-R)^2}{R}+\frac{(1-\rho)^2}{\rho}\right\}\right] \qquad \ldots\ldots(4)$$

as a working formula.

The accuracy attained may be illustrated by a numerical example. Suppose the true values are

$$\epsilon = 0{\cdot}05, \quad \tau = 0{\cdot}08, \quad \rho = 0{\cdot}90.$$

Calculation by (2) gives

$$R = 0{\cdot}5851, \quad \frac{R}{\rho} = 0{\cdot}6501, \quad \frac{(1-R)^2}{R} + \frac{(1-\rho)^2}{\rho} = 0{\cdot}3053.$$

Substituting in (4) and using the correct value of ϵ gives

$$\tau^2 = 0{\cdot}6402,$$

a very satisfactory result. The magnitude of the difference between τ^2 and R/ρ indicates that ϵ should be known with an accuracy of a few units in the third decimal place. If the surfaces are bloomed the term in ϵ should be negligible.

ACKNOWLEDGMENT

The work described above has been carried out as part of the research programmes of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCE

Marechal, A., 1944. " La mesure directe du facteur d'absorption des verres optiques." *Paris*: *Communications des Laboratoires de l'Institut d'Optique.*, No. 14.

DEMONSTRATION OF A WATER-JET ANALOGUE OF THE REFLECTION KLYSTRON

By W. J. SCOTT,
B.T.H. Research Laboratory, Rugby

Demonstrated 8 *June* 1945

The reflection klystron has a rhumbatron-type cavity resonator with two closely spaced grids through which an initially unmodulated electron beam passes. The field between the grids alternates, accelerating or retarding the electrons, which pass on into a retarding field which returns them through the grids. The accelerated electrons go further than the retarded ones and so take longer to return. With proper adjustment of beam voltage and reflection field, the resonator receives over one half-cycle more energy than it gives up.

It is often taught that electric current is like water flowing through a pipe, and that a triode valve is like a water-valve. Extending the analogy (Scott, 1944), an electron beam can be represented by a water jet and an electric field by the earth's gravitational field. On this basis a working model equivalent to a reflection oscillator has been made, using a balanced horizontal arm A mounted on two adjustable leaf-spring hinges B as the tunable "resonator" (figure 1). On one end of the arm, a close-mesh gauze "catcher" grid C is fixed, and behind it, but 10 in. distant, a nearly vertical nozzle D which is fed with water through the hollow arm from a flexible rubber tube E. Figures 1 to 4 show the model in oscillation. The photographs were taken using a B.T.H. electrical discharge photo-flash lamp set to give a flash duration of 100 micro-seconds.

Starting from rest, the random motion of the water as it breaks into globules produces sufficient agitation (corresponding to electron agitation) to initiate oscillation provided the jet velocity is correct. In action the jet is modulated in velocity by the vertical simple-harmonic motion of the nozzle. The water accelerated during the second half of the up-stroke and that retarded on the first half of the down-stroke bunch together and arrive back to hit and propel the gauze grid on its downward stroke. Thus over one half-cycle more energy is given to the system than is absorbed by friction, windage, and by water impinging on the upward stroke.

Figure 1 shows a compacted bunch, FG, half-way back, and at the top a still widely spread "future" bunch, HJ. In figure 2 the beam is moving down, and is about to be hit by the now horizontally disposed bunch KL. At the top the next bunch, MN, is compacting. In figure 3 the bunch OP is splashing through the grid. In figure 4 the grid is moving up again, the next bunch, QR, forms a loop inclined at 45°, and a rising part, ST, of the jet, inclined at 10°, will form another. In figure 2 two well-formed bunches and another partly formed one are in the air.

Oscillations can be maintained by flight times corresponding to approximately 1¼, 2¼, 3¼, etc., cycles, instead of the ¾, 1¼, 2½, etc., cycles, which are required by the electron tube. The slight increase or decrease of frequency, caused by slightly lowering or increasing the jet velocity, so that "bunches" arrive sooner or later at the catcher, corresponds strictly to the electronic tuning of reflex tubes (Pierce, 1945). If the jet velocity is definitely incorrect, oscillations will not build up or even maintain themselves.

REFERENCES

Brainerd, Koehler, Reich and Woodruff, 1942. *Ultra High Frequency Techniques* (Chapman and Hall), p. 339.

British Patent 545,146.

Hahn, W. C. and Metcalf, G. F., 1939. *Proc. Inst. Radio Engrs.*, **27**, 109.

Pierce, J. R., 1945. *Proc. Inst. Radio Engrs.*, **33**, 112.

Scott, W. J., 1944. *J. Instn. Elect. Engrs.*, **91**, 143.

Varian, R. H. and S. F., 1939. *J. Appl. Phys.*, **10**, 140.

PHYSICAL SOC. VOL. 58, PT. 4 (W. J. SCOTT)

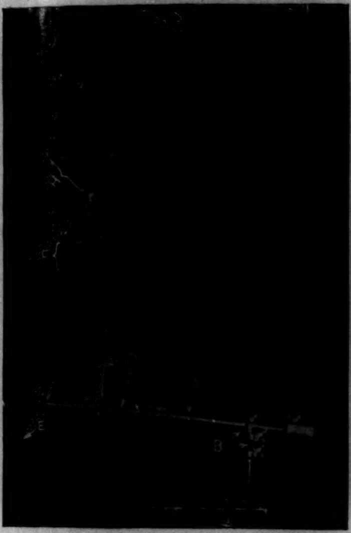

Figure 1.

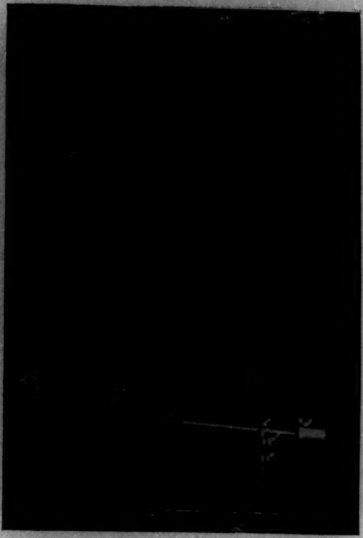

Figure 2.

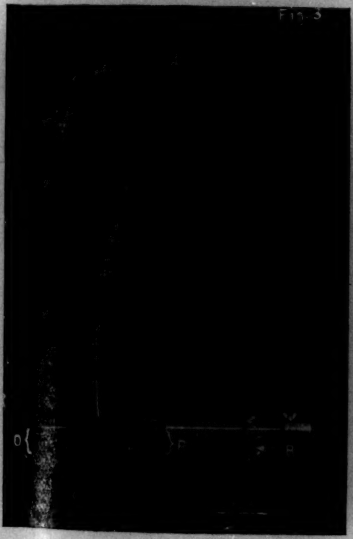

Figure 3.

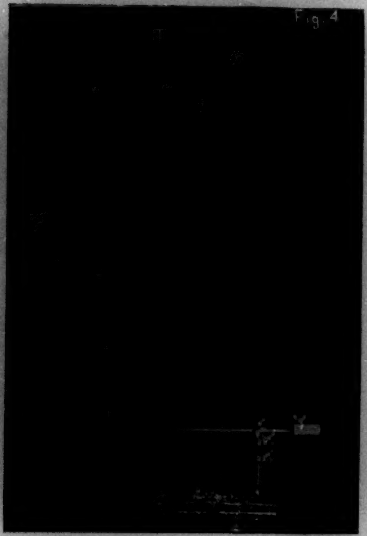

Figure 4.

To face page 476

NOTE ON THE MEASUREMENT OF SPECTRAL DISTRIBUTIONS WHEN LINES AND CONTINUA ARE PRESENT TOGETHER

By MARY P. LORD,
Philips' Lamps, Ltd.
Now at Imperial College, London

MS. received 11 *February* 1946

ABSTRACT. The problem of evaluating the intensity of a line spectrum relative to that of a continuum is solved for the specific case of a double monochromator (the dispersions in the two halves of which are in the same direction) used in conjunction with a selective detector. The form of a function of the slit-widths and the dispersion required for the solution is derived analytically: a geometrical treatment given by Cittert is criticized.

---

§ 1. INTRODUCTION

As the radiation from the majority of fluorescent lamps consists of lines (mainly those of mercury) superimposed on phosphor continua, the problem of evaluating the intensity of a line spectrum relative to that of a continuum is one of practical importance. Despite this it appears to have received little attention.

§ 2. SOLUTION IN TERMS OF AN UNKNOWN FUNCTION OF THE SLIT-WIDTHS AND THE DISPERSION

To avoid circumlocution, the solution of the problem is given for the specific case of a double monochromator, the dispersions in the two halves of which are in the same direction (Perry, 1938) used in conjunction with a selective detector. It is further supposed that for each wave-length the energy of continuum per unit of wave-length (at the entrance slit) required to produce a given detector response is known. Solutions for other cases can easily be deduced.

Consider a source which radiates in unit time an amount of energy ϵ_λ as a line of wave-length λ, and a continuum whose energy in the range λ to $\lambda + d\lambda$ is $E_\lambda d\lambda$. Let the widths of the entrance, middle and exit slits of the monochromator be respectively s_1, s_2 and s_3; they are assumed to be sufficiently wide for diffraction effects to be negligible. Then, as the magnification produced by the monochromator at each slit is unity, the energy of line of wave-length λ emerging per unit time from the exit slit when the monochromator is set for this wave-length is $\kappa_\lambda s \epsilon_\lambda$, where s is the smallest of s_1, s_2 and s_3, and κ_λ is a factor depending on the aperture of the monochromator and the absorption of its optical system. If the slits are sufficiently narrow for E_λ to be sensibly constant over the range of wave-lengths embraced by them (which should of course be the case) the energy of continuum emerging can be expressed as $\kappa_\lambda E_\lambda . f(s_1, s_2, s_3, \phi_\lambda)$, where f is some function of s_1, s_2, s_3 and ϕ_λ; ϕ_λ is the reciprocal of the linear dispersion of either half of the monochromator at wave-length λ. The form of this function is given in § 3.

The indication of the detector when the monochromator is set at wave-length λ is of course produced by the line and continuum together. Let E_λ' (obtained from the calibration curve) be the energy of continuum per unit of wave-length which would correspond to this detector reading. Then the part of E_λ' to which the line gives rise is $E_\lambda' - E_\lambda$, where E_λ is the part of E_λ' to which the continuum gives rise, which can be obtained by interpolation between continuum values on either side of the line. Since a given detector indication at a given wave-length always corresponds to the same amount of energy incident on the detector, the energy from the line which strikes the detector is

$$\kappa_\lambda(E_\lambda' - E_\lambda) . f(s_1, s_2, s_3, \phi_\lambda).$$

Hence $$\kappa_\lambda s \epsilon_\lambda = \kappa_\lambda(E_\lambda' - E_\lambda) . f(s_1, s_2, s_3, \phi_\lambda),$$

i.e. $$\epsilon_\lambda = (E_\lambda' - E_\lambda) . f(s_1, s_2, s_3, \phi_\lambda)/s,$$

and thus the energy of the line relative to that of the continuum in the emitted light may be obtained.

In this treatment it has been assumed that there are no lines whose wave-lengths are sufficiently close to λ for them to pass the exit slit (in whole or part) when the monochromator is set at wave-length λ. Measurements can, however, still be made even if this assumption is not justified. It is only necessary to calculate the wave-length setting of the monochromator for the part of the line which is not overlapped, make the observation in this position, and apply the formula given above with the width of the part of the line observed substituted for s. The accuracy of this procedure is, of course, limited by the mechanism of the monochromator.

§ 3. ANALYTIC DERIVATION OF THE FORM OF THE FUNCTION

Using a geometrical method, Cittert (1926) has obtained expressions for f, but his treatment of the range of validity of the solution is not entirely satisfactory, and the dependence on the dispersion is not mentioned. These criticisms will be amplified after the analytical treatment has been given.

It is first necessary to obtain an expression for the intensity of radiation passing the second slit. There are obviously three cases to consider: $s_1 \gtreqless s_2$. Since the width of the image of any wave-length in the plane of the middle slit is s_1 in the first case ($s_1 > s_2$), the energy of the radiation in the wave-length range dx surrounding any wave-length x in the range $\lambda - \frac{s_1 - s_2}{2}\phi_\lambda$ to $\lambda + \frac{s_1 - s_2}{2}\phi_\lambda$ will be uniform (namely $k_\lambda E_\lambda s_2 \, dx$), but the energy in the band surrounding wave-lengths outside this range will steadily fall until it reaches zero at wave-lengths $\lambda \pm \frac{s_1 + s_2}{2}\phi_\lambda$. Thus the intensity in the range dx surrounding any wave-length x in the regions of varying intensity is

$$k_\lambda E_\lambda \left[\frac{s_1 + s_2}{2} - \frac{|\lambda - x|}{\phi_\lambda}\right] dx,$$

where k_λ is analogous to κ_λ. Hence in the case $s_1 > s_2$ the intensity passing the middle slit is

$$\int_{\lambda - \frac{s_1 - s_2}{2}\phi_\lambda}^{\lambda + \frac{s_1 - s_2}{2}\phi_\lambda} k_\lambda E_\lambda s_2 \, dx + 2\int_{\lambda - \frac{s_1 + s_2}{2}\phi_\lambda}^{\lambda - \frac{s_1 - s_2}{2}\phi_\lambda} k_\lambda E \left[\frac{s_1 + s_2}{2} - \frac{\lambda - x}{\phi}\right] dx,$$

which comes to $k_\lambda \phi_\lambda E_\lambda s_1 s_2$.

The expression for the intensity in the case $s_1 < s_2$ differs only by interchange of s_1 and s_2, and hence is also equal to $k_\lambda \phi_\lambda E_\lambda s_1 s_2$. This expression holds, again, when $s_1 = s_2$.

The effect of the third slit must now be considered. There are obviously three physically distinct cases:

1. None of the radiation is cut off at the third slit.
2. The third slit only passes wave-lengths which arrive in its plane with the same intensity.
3. The third slit cuts off in the region of diminishing intensity. The conditions satisfied by the slit-widths for each of these three cases can easily be written down since the width of the image of any wave-length in the plane of the third slit is the width of the part of the image of this wave-length which passes the second slit. Using the same principle, expressions for the intensity emerging in each of the three cases can also be obtained.

Thus for $s_1 \gtreqless s_2$ the condition for the first case is $s_3 \geqslant 2s_2 + s_1$; the intensity emerging is obviously $\kappa_\lambda E_\lambda \phi_\lambda s_1 s_2$, i.e. $f \equiv \phi_\lambda s_1 s_2$.

For the second case, if (*a*) $s_1 > s_2$ the condition is $s_3 \leqslant s_1 - 2s_2$; this is again effectively a two-slit case (s_2 and s_3), and so for $s_2 \gtreqless s_3$ the intensity emerging is $\kappa_\lambda E_\lambda \phi_\lambda s_2 s_3$, i.e. $f \equiv \phi_\lambda s_2 s_3$. The condition for the second case if (*b*) $s_1 < s_2$ is $s_3 \leqslant 2s_2 - 3s_1$. This is effectively another two-slit case (s_1 and s_3), the dispersion being twice that of the previous one (since the slits in this case are separated by a double optical system). Hence the intensity emerging for $s_1 \gtreqless s_3$ is $\frac{1}{2}\kappa_\lambda E_\lambda \phi_\lambda s_1 s_3$, i.e. $f \equiv \frac{1}{2}\phi_\lambda s_1 s_3$. This case does not occur if $\frac{2}{3} s_2 \leqslant s_1 \leqslant 2s_2$.

For the third case, if (*a*) $s_1 > s_2$ the condition is $2s_2 + s_1 > s_3 > s_1 - 2s_2$; the wave-lengths at which the slit cuts off completely are $\lambda \pm \dfrac{s_1 + s_3}{4}\phi_\lambda$, and the wave-lengths between which there is no cutting off are $\lambda \pm \dfrac{s_3 - s_2}{2}\phi_\lambda$. Hence, by reasoning analogous to that given in the two-slit case, the expression for the emerging intensity is

$$\kappa_\lambda E_\lambda \left[2\int_{\lambda - \frac{s_1+s_3}{4}\phi_\lambda}^{\lambda - \frac{s_3-s_2}{2}\phi_\lambda} \left(\frac{s_1+s_3}{2} - \frac{2(\lambda - x)}{\phi_\lambda}\right) dx \right.$$
$$\left. + 2\int_{\lambda - \frac{s_3-s_2}{2}\phi_\lambda}^{\lambda - \frac{s_1-s_2}{2}\phi_\lambda} \left(\frac{s_1+s_3}{2} - \frac{\lambda - x}{\phi_\lambda}\right) dx + \int_{\lambda - \frac{s_1-s_2}{2}\phi_\lambda}^{\lambda + \frac{s_1-s_2}{2}\phi_\lambda} s_2\, dx \right],$$

which comes to

$$\kappa_\lambda E_\lambda \phi_\lambda \left[\frac{s_2}{2}(s_1 + s_3 - s_2) - \frac{(s_1 - s_3)^2}{8}\right],$$

i.e.

$$f \equiv \phi_\lambda \left[\frac{s_2}{2}(s_1 + s_3 - s_2) - \frac{(s_1 - s_3)^2}{8}\right].$$

If (*b*) $s_1 < s_2$, the condition is $2s_2 + s_1 > s_3 > 2s_2 - 3s_1$, and the expression for the intensity is

$$\kappa_\lambda E_\lambda \left[2\int_{\lambda - \frac{s_1+s_3}{4}\phi_\lambda}^{\lambda - \frac{s_3-s_2}{2}\phi_\lambda} \left(\frac{s_1+s_3}{2} - \frac{2(\lambda - x)}{\phi_\lambda}\right) dx \right.$$
$$\left. + 2\int_{\lambda - \frac{s_3-s_2}{2}\phi_\lambda}^{\lambda - \frac{s_2-s_1}{2}\phi_\lambda} \left(\frac{s_1+s_2}{2} - \frac{\lambda - x}{\phi}\right) dx + \int_{\lambda - \frac{s_2-s_1}{2}\phi_\lambda}^{\lambda + \frac{s_2-s_1}{2}\phi_\lambda} s_1\, dx \right],$$

which again comes to

$$\kappa_\lambda E_\lambda \phi_\lambda \left[\frac{s_2}{2}(s_1+s_3-s_2) - \frac{(s_1-s_3)^2}{8}\right].$$

The conditions for, and intensity emerging in, this case if (*c*) $s_1 = s_2$ are given by substituting this relationship in either of the above conditions and expressions, respectively.

To summarize, the conditions for the various cases are therefore:

1 $s_1 \gtreqless s_2,\quad s_3 \geqslant 2s_2 + s_1.$

2 (*a*) $s_2 \gtreqless s_3,\quad s_3 \leqslant s_1 - 2s_2;$

(*b*) $s_1 \gtreqless s_3,\quad s_3 \leqslant 2s_2 - 3s_1.$

3 (*a*) $2s_2 + s_1 > s_3 > s_1 - 2s_2;$

(*b*) $2s_2 + s_1 > s_3 > 2s_2 - 3s_1;$

(*c*) $3s_2 = 3s_1 > s_3.$

In his treatment of the problem, Cittert (1926) distinguishes four cases (according to the intensity emerging) and gives the conditions for them as:

A $2s_2 > s_1 + s_3.$
B $2s_2 < s_3 + s_1$ and $s_3 > s_1 + 2s_2.$
C $2s_2 < s_3 + s_1$ and $s_3 < s_1 + 2s_2$ and $s_1 < s_3 + 2s_2.$
D $2s_2 < s_1 + s_3$ and $s_1 > s_3 + 2s_2.$

There are several objections to this. Thus it is not pointed out that case B (corresponding to 1) is significant for $s_1 \gtreqless s_2$; the second condition of case B implies the first. A similar objection applies to D (which corresponds to 2 (*a*)); moreover, it is not stated that D is significant for $s_2 \gtreqless s_3$; that A (corresponding to 2 (*b*)) is significant for $s_1 \gtreqless s_3$ is not mentioned either—the condition given is inaccurate. The statement of the conditions for case C is misleading: if $s_1 > s_2$ (corresponding to 3 (*a*)), the first, which is inaccurate, is irrelevant, and if $s_1 < s_2$ (corresponding to 3 (*b*)), then the last one is irrelevant; mention of 3 (*c*) is omitted. This method of presentation, moreover, somewhat obscures the physical significance of the conditions.

ACKNOWLEDGMENTS

The author is indebted to Mr. J. A. M. van Moll and to the Directors of Philips' Lamps, Ltd., for permission to publish this note.

REFERENCES

Perry, 1938. *Proc. Phys. Soc.*, **50**, 265.
van Cittert, 1926. *Z. Instrumkde*, **46**, 557.

THE TRANSIENT FLOW OF HEAT THROUGH A TWO-LAYER WALL

By MIRIAM V. GRIFFITH and G. K. HORTON,*
B.E.A.I.R.A., Greenford, Middlesex
* Now at Imperial College of Science and Technology

MS. received 23 *March* 1946

ABSTRACT. The sudden introduction of a high temperature source of heat into a room is idealized by supposing that the walls absorb a constant uniform flux of heat. Formulae for the temperature distribution after various times are derived on the hypothesis that a wall is of infinite extent and comprises a uniform finite layer superimposed on an indefinitely thick base. The factors which determine the rate of rise of the wall-surface temperature are particularly studied.

---

§ 1. INTRODUCTION

THE report gives a mathematical treatment of the problem of the temperature distribution in a non-homogeneous wall of two components under transient conditions when the flow of heat into the surface of the wall is constant. This is of particular interest when considering the rise of temperature of insulated wall surfaces in a domestic room, heated by radiation from a high-temperature source. The particular application envisaged is when such a room is adapted, by means of suitable wall coverings, to promote conditions of comfort in as short a time as possible after applying the source of heat.

From an economic point of view, it is desirable to know the thickness of thermal insulation required, and the effect on the temperature distribution of the thermal properties of the insulation material. By means of certain assumptions, which have been justified experimentally, it is possible to form an equation from which curves can be constructed for practical use.

§ 2. MATHEMATICAL FORMULAE

Assume first that the area of the wall is great, so that the heat is conducted perpendicularly to its surface, which is taken as the plane $x=0$. The wall consists of two components, one of thickness α, extending from the plane $x=0$ to $x=\alpha$, while the second component extends from $x=\alpha$ to $x=\infty$ (under practical conditions, and for the times and materials under consideration, this conception introduces no error).

The temperature $U(x, t)$ satisfies the following conditions :

$$\frac{dU}{dt}=D\frac{d^2U}{dx^2}\begin{array}{lll} t>0, & \alpha\geqslant x\geqslant 0; & D=D_1 \\ t>0, & \infty\geqslant x\geqslant \alpha; & D=D_2, \end{array}\Bigg\} \qquad \ldots\ldots(1)$$

$$\operatorname*{Lt}_{t=0} U_1(x,t)=0, \qquad \ldots\ldots(2)$$

$$\operatorname*{Lt}_{x=\alpha} U_1(x,t)=\operatorname*{Lt}_{x=\alpha} U_2(x,t), \qquad \ldots\ldots(3)$$

$$\operatorname*{Lt}_{x=\alpha} k_1 \frac{dU_1}{dx} = \operatorname*{Lt}_{x=\alpha} k_2 \frac{dU_2}{dx}, \qquad \ldots\ldots(4)$$

$$\operatorname*{Lt}_{x=0} \left(-k_1 \frac{dU_1}{dx} \right) = H \text{ for } 0<t<\infty, \qquad \ldots\ldots(5)$$

$$\operatorname*{Lt}_{x=\infty} U_2(x,t) = 0 \text{ for } 0<t<\infty, \qquad \ldots\ldots(6)$$

where subscript 1 refers to top layer 1, subscript 2 refers to base layer 2, and

x = distance perpendicular to the face of the wall,
t = time,
D = thermal diffusivity = k/S.
k = thermal conductivity,
H = constant flux of heat perpendicular to the wall surface,
S = volume specific heat (specific heat × density).

Let

$$\overline{U} = \int_0^\infty U e^{-pt}\, dt, \text{ the Laplace transform of } U;$$

then

$$\frac{d\overline{U}}{dt} = p\overline{U} - U(0);$$

hence

$$\frac{d^2\overline{U}}{dx^2} - \frac{p}{D}\,\overline{U} = 0 \text{ since } U_{\lambda}(0) = 0. \qquad \ldots\ldots(7)$$

Thus

$$\overline{U}_1 = A_1 e^{-\sqrt{p/D_1}\,x} + B_1 e^{\sqrt{p/D_1}\,x}, \qquad \ldots\ldots(8)$$

$$\overline{U}_2 = A_2 e^{-\sqrt{p/D_2}\,x}. \qquad \ldots\ldots(9)$$

The boundary conditions give

$$-A_1\sqrt{p/D_1} + B_1\sqrt{p/D_1} = -H/k_1 p \text{ at } x=0; \qquad \ldots\ldots(10)$$

also

$$\frac{k_1}{k_2} = \frac{(-A_2/\sqrt{D_2})e^{-\alpha\sqrt{p/D_2}} + (B_2/\sqrt{D_2})e^{\alpha\sqrt{p/D_2}}}{(-A_1/\sqrt{D_2})e^{-\alpha\sqrt{p/D_1}} + (B_1/\sqrt{D_1})e^{\alpha\sqrt{p/D_1}}} \qquad \ldots\ldots(11)$$

and

$$A_1 e^{-\alpha\sqrt{p/D_1}} + B_1 e^{\alpha\sqrt{p/D_1}} = A_2 e^{-\alpha\sqrt{p/D_2}} + B_2 e^{\alpha\sqrt{p/D_2}}. \qquad \ldots\ldots(12)$$

These lead to

$$B_1 = -\frac{\dfrac{H\sqrt{D_1}}{k_1 p\sqrt{p}}\left[e^{-\alpha\sqrt{p/D_1}} - \dfrac{k_2}{k_1}\sqrt{\dfrac{D_1}{D_2}}\, e^{-\alpha\sqrt{p/D_1}} \right]}{(e^{-\alpha\sqrt{p/D_1}} - e^{\alpha\sqrt{p/D_1}}) - \dfrac{k_2}{k_1}\sqrt{\dfrac{D_1}{D_2}}(e^{-\alpha\sqrt{p/D_1}} + e^{\alpha\sqrt{p/D_1}})}$$

$$= \frac{-H\sqrt{D_1}}{\gamma k_1 p\sqrt{p}} \frac{1}{e^{2\alpha\sqrt{p/D_1}}} \left(\frac{1}{1+\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}}} \right), \qquad \ldots\ldots(13)$$

where

$$\gamma = \frac{k_2 S_2 + k_1 S_1}{k_2 S_2 - k_1 S_1}.$$

Similarly

$$A_2=(2H/p)\sqrt{D_1D_2/p}\ \text{exp.}\ \alpha\sqrt{p}(D_1^{-\frac{1}{2}}+D_2^{-\frac{1}{2}})\{\lambda/(1+\gamma e^{2\alpha\sqrt{p/D_1}})\}, \qquad \ldots\ldots(14)$$

where

$$\lambda=(k_2\sqrt{D_1}-k_1\sqrt{D_2})^{-1}.$$

Hence

$$\bar{U}_1=\frac{H\sqrt{D_1}}{k_1p\sqrt{p}}\left[1-\frac{1}{\gamma e^{2\alpha\sqrt{p/D_1}}}\left(\frac{1}{1+\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}}}\right)\right]e^{-x\sqrt{p/D_1}}$$

$$-\frac{H\sqrt{D_1}}{k_1p\sqrt{p}}\frac{1}{e^{2\alpha\sqrt{p/D_1}}}\left(\frac{1}{1+\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}}}\right)e^{x\sqrt{p/D_1}} \qquad \ldots\ldots(15)$$

and

$$\bar{U}_2=\frac{2H\sqrt{D_1D_2}\lambda}{\gamma p\sqrt{p}}\left(\frac{1}{1+\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}}}\right)e^{\sqrt{p}\left[\alpha\left(\frac{1}{\sqrt{D_1}}+\frac{1}{\sqrt{D_2}}\right)-\frac{x}{\sqrt{D_2}}\right]}.$$

$$\ldots\ldots(16)$$

Now $\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}}$ is essentially less than unity, so $(1+\gamma^{-1}e^{-2\alpha\sqrt{p/D_1}})^{-1}$ can be expanded as a power series. If so, then, as functions of p, $\bar{U}_1$ and $\bar{U}_2$ will be the sum of terms of the form $(A/p\sqrt{p})e^{-B\sqrt{p}}$. The inverse transforms of such functions of p are known, so that when the expansions have been derived, the inverse transforms U_1 and U_2 may be written down as

$$U_1=\frac{H}{k_1}\left\{\left[2\sqrt{\frac{D_1t}{\pi}}\,e^{-x^2/4D_1t}-x\left(1-\operatorname{erf}\frac{x}{2\sqrt{D_1t}}\right)\right]\right.$$

$$-\frac{1}{\gamma}\sum_{n=0}^{n=\infty}\left(-\frac{1}{\gamma}\right)^n\left[2\sqrt{\frac{D_1t}{\pi}}\left(e^{-\{x+2\alpha(n+1)\}^2/4D_1t}+e^{-\{x-2\alpha(n+1)\}^2/4D_1t}\right)\right.$$

$$-\{x+2\alpha(n+1)\}\left\{1-\operatorname{erf}\left(\frac{x+2\alpha(n+1)}{2\sqrt{D_1t}}\right)\right\}$$

$$\left.\left.+\{x-2\alpha(n+1)\}\left\{1-\operatorname{erf}\left(-\frac{x-2\alpha(n+1)}{2\sqrt{D_1t}}\right)\right\}\right]\right\}; \qquad \ldots\ldots(17)$$

$$U_2=\frac{2H\lambda\sqrt{D_1}}{\gamma}\sum_{n=0}^{n=\infty}\left(-\frac{1}{\gamma}\right)^n\left\{2\sqrt{\frac{D_2t}{\pi}}\,e^{-\{x-\alpha\{1-\sqrt{D_2/D_1}(2n+1)\}\}^2/4D_2t}\right.$$

$$\left.-[x-\alpha\{1-\sqrt{D_2/D_1}(2n+1)\}]\left(1-\operatorname{erf}\frac{x-\alpha\{1-\sqrt{D_2/D_1}(2n+1)\}}{2\sqrt{D_2t}}\right)\right\}.$$

$$\ldots\ldots(18)$$

U_1 gives the temperature at any point x in layer 1 of thickness α at any time t from the introduction of the source of heat, while U_2 gives the temperature in the semi-infinite layer 2 of the wall.

The surface temperature U_0 is given by putting $x=0$ in the expression for U_1, viz.,

$$U_0=\frac{H}{k_1}\left[2\sqrt{\frac{D_1t}{\pi}}-\frac{4}{\gamma}\sum_{n=0}^{n=\infty}\left(-\frac{1}{\gamma}\right)^n\left\{\sqrt{\frac{D_1t}{\pi}}\,e^{-\alpha^2\frac{(n+1)^2}{tD_1}}\right.\right.$$

$$\left.\left.-\alpha(n+1)\left(1-\operatorname{erf}\frac{\alpha(n+1)}{\sqrt{D_1t}}\right)\right\}\right]. \qquad \ldots\ldots(19)$$

The temperature at the interface, U_S, is given by U_1 when $x=\alpha$:

$$U_S=\frac{H}{k_1}\sum_{n=0}^{n=\infty}\left(-\frac{1}{\gamma}\right)^n\left(1-\frac{1}{\gamma}\right)\left[2\sqrt{\frac{D_1t}{\pi}}\,e^{-\left(\frac{\alpha(2n+1)}{2\sqrt{D_1t}}\right)^2}-\alpha(2n+1)\left(1-\operatorname{erf}\frac{\alpha(2n+1)}{2\sqrt{D_1t}}\right)\right]. \qquad \ldots\ldots(20)$$

§ 3. APPLICATION OF RESULTS

Representative values of k_1S_1, k_2S_2 for a practical range of materials have been used to calculate the variation of temperature rise of the insulation surface for various times, with the results illustrated in figure 1. This shows the behaviour of the best insulating materials, both as regards high thermal resistivity and low volume specific heat, for which data are available, when used as a wall covering on brick of 5% water content.

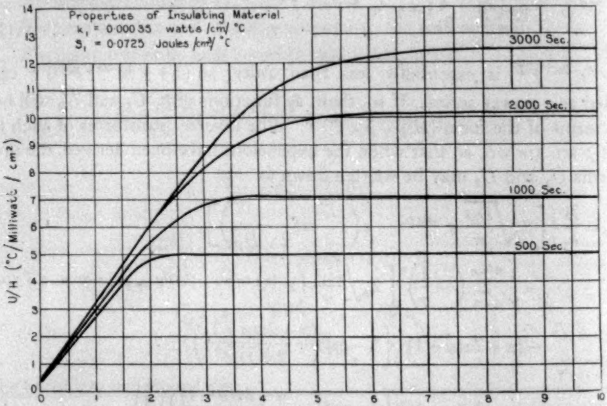

Figure 1. Variation of surface temperature with thickness of insulation material on brick.

It will be observed that for each time of heating there exists a thickness after which the wall behaves as if it were of infinite thickness, and the surface temperature rise corresponding to this is, therefore, the maximum obtainable with that material for the time of heating concerned. Provided this thickness is attained, the base layer 2 may be ignored and the surface temperature will follow the law for a homogeneous wall, viz.,

$$U_0=2H\sqrt{(t/S_1k_1\pi)}=(2H/k_1)\sqrt{(D_1t/\pi)}. \qquad \ldots\ldots(21)$$

Now in figure 2 has been plotted the thickness α for which the lower layer affects the surface temperature by 10%, and thicknesses exceeding this may be taken as effectively infinite in the application envisaged. This critical thickness is given as a series of curves against D_1t for various values of k_1S_1, assuming the lower layer to be brick. For thicknesses greater than the critical value of α, equation (21) applies, so that if a certain value of U_0/H is assumed, the scale of D_1t can be transformed to a scale of k_1, which has been inserted for a value of U_0/H of 1° c. per milliwatt per cm². This means that, if it is desired to know the time to reach a temperature rise of 1° c. per milliwatt per cm², then by noting the value of the thermal conductivity of the wall covering on the scale of k_1, D_1t is found. If the volume specific heat is known, then D_1 is known and t, the time required, is obtained, while by reading α on the curve corresponding to the appropriate value of k_1S_1, the minimum thickness of wall covering can be found. For other values of U_0/H, figure 3 can be used, which gives U_0/H as a function of the time t from equation (21) for different values of k_1S_1. If the properties of the wall covering are known, the time of heating, temperature rise and intensity of the source of heat are given by the curve with the appropriate value of k_1S_1, while figure 2 gives the minimum thickness of covering.

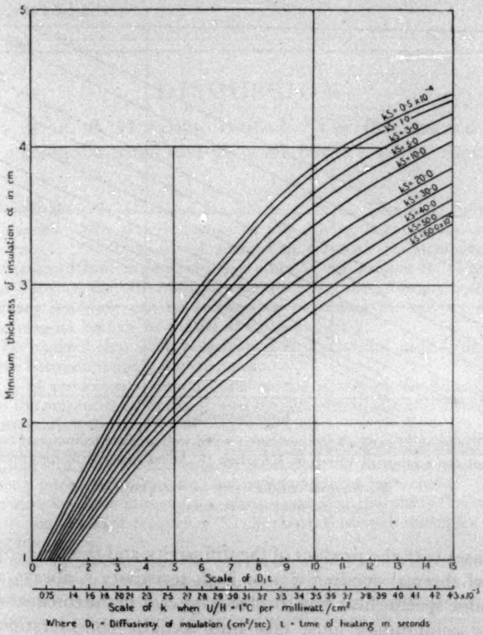

Figure 2. Variation with Dt and kS of minimum thickness of insulation on a brick wall to give maximum surface temperatures.

Conversely, if the desired temperature rise and intensity of source (giving the intensity of radiation on the wall surface) are known, then the time to attain this temperature can be found from figure 3 for a series of values of k_1S_1. If any particular material is chosen to satisfy a chosen value of k_1S_1, then D_1 is also known together with D_1t, which, from figure 2, will then give the minimum thickness, or else, the thickness having been decided, D_1t can be found from figure 2 for any value of k_1S_1, which again gives the thermal conductivity and volume specific heat of the material required.

As is well known, the thermally insulating properties of a material for transient heating of this type depend ideally only on the product of its thermal conductivity and volume specific heat, decreasing as the latter increases, but the thickness which is required so that, for the given time of heating, the base material does not matter, increases with the product of the diffusivity and the time involved. Thus a low value of thermal conductivity or high resistivity is always advantageous, but a low volume specific heat may involve some cost in the thickness of insulation. However, the compromise must take into account considerations other than scientific, so that this aspect cannot further be pursued here.

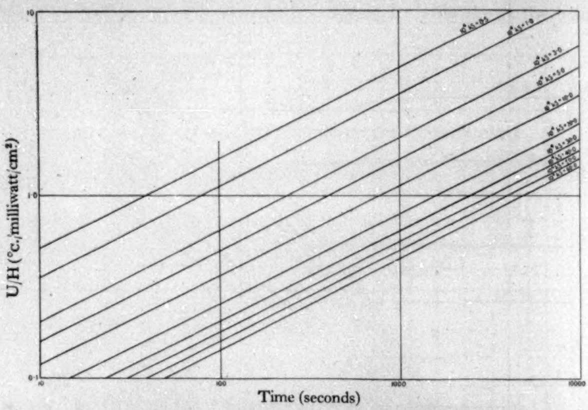

Figure 3. Variation of ratio of temperature rise of wall surface to incident radiation with time for a range of values of kS when thickness exceeds minimum (α).

k = thermal conductivity in watts/cm./°c.
S = specific heat in watts/cm³/°c.

§ 4. CONCLUSIONS

(1) With a constant flux of heat a two-layer wall will behave effectively as a single layer, provided the time of heating is short enough in relation to the thickness of the surface layer and its properties. Formulae are derived for the transient temperature distributions.

(2) This critical thickness varies mainly by increasing with increase of the product of diffusivity of the surface layer and time, for a given base material, and decreases to a less extent with increase of the product of thermal conductivity and volume specific heat,

(3) While for a homogeneous wall the thermally insulating properties in transient heating require a low value of the product of volume specific heat and thermal conductivity, yet a low volume specific heat may increase the thickness of the material required to obviate the effect of a base material of conducting properties on which the first material may be applied.

§ 5. ACKNOWLEDGMENTS

The authors are indebted to the Director of the British Electrical and Allied Industries Research Association for permission to publish this paper, and for his helpful criticism and advice.

DISCUSSION

on paper by R. F. S. Hearmon entitled "The fundamental frequency of vibration of rectangular wood and plywood plates" (*Proc. Phys. Soc.*, **58**, 78, (1946)).

Dr. Kantorowicz. It is stated in § 3 that "... The most powerful response in all cases occurred when the frequency of the driving current was one-half the true resonant frequency ..." This is not surprising, for with the arrangement shown on page 83 the armature A is attracted twice per cycle by the magnet B. There is, therefore, strict resonance between the vibration exciting force and the vibrating body.

This frequency doubling can be avoided by polarizing the magnet either by means of a permanent magnet core or by means of a polarizing D.C.

It should be realized that such polarization is a powerful means for magnifying the alternating force between magnet and armature.

The force F of attraction is proportional to the square of the flux, which in turn is proportional to the sum of the polarizing and the alternating currents i and δi respectively; therefore the alternating force is $F+\delta F=a(i+\delta i)^2$.

If the static force, $F=ai^2$, is subtracted it is seen that $\delta F=a(2i\delta i+\delta i^2)$.

Even if δi follows a sine law the force of attraction will contain a term in $\sin^2$. Consequently it is not a pure harmonic force, but contains higher harmonics.

This is the simple explanation of the observation that the plates " respond to integral sub-multiples of the resonant frequency ". In short, those sub-multiples have harmonics at the resonant frequency.

It is noted that the author feeds the voltage developed across the magnet to an oscilloscope. It is recommended that instead he should deflect the electron beam by a signal in phase with the current and, therefore, in phase with the exciting force, for in this case he can easily observe whether the velocity of plate motion is in phase with the exciting force. (His pick-up, $C\times D$, is velocity-sensitive.)

This condition of exciting force and velocity of vibration being in phase is a very sharp criterion of resonance and much more easily recognized than maximum amplitude.

With improved technique I expect the author will be able to supplement the information of resonant frequency by the probably even more important one of the damping capacities.

Author's reply. I am grateful to Dr. Kantorowicz for his comments, and hope to try out the practical suggestions in any future work on vibration problems. The arrangement he proposes for avoiding frequency doubling and for magnifying the alternating driving force has not, to my knowledge, been used in experiments on the vibration of plates, but I have used a similar arrangement, due to Ide (1935) (*Rev. Sci. Instrum.*, **6**, 296), to excite longitudinal vibrations in wooden specimens. The most powerful response then occurred with the frequency of the driving current equal to the resonant frequency, in agreement with Dr. Kantorowicz's prediction. The existence of the frequency doubling was, of course, recognized and allowed for in my paper, and the results are, therefore, not affected by it.

REVIEWS OF BOOKS

Mathematical Theory of Elasticity, by I. S. Sokolnikoff. Pp. 394. (Brown University: Summer Session for Advanced Instruction and Research in Mechanics, June 23rd–September 13th, 1941.)

The first reaction to this course of lecture notes must be a feeling of admiration for the American student, the second is likely to be one of pity. Previously equipped with a good working knowledge of applied mathematics, the summer student (and how warm is it in Providence in summer ?) is expected to imbibe six large doses of an elixir compounded of the essences of Love and Timoshenko and laced with rarer precious oils from far-off Tartary. All students who may stay the course will certainly be finished elasticians ; some others may be tempted to insert the conjunction " as ".

Assuming that the title-page is to be disregarded and that the draught is not to be swallowed in such indecent haste, it seems odd that so important a work as this should be given no more than cyclostyled circulation. Even in the course of review the 394 pages have become hopelessly dog-eared, and the reviewer's present preoccupation is to find some way to preserve the tattered remnants in some sort of usable form. For, despite its production and title, this presentation of the mathematical theory of elasticity takes its proper place as a standard work of reference, and the very lowest appraisal must rank it as a valuable supplement to Love and Timoshenko.

Following the King of Hearts' advice, the work starts at the beginning, and the first page clearly states those limitations of the scope of the theory which are too often overlooked. The remainder of the King of Hearts' advice is less well followed, and the scope of the course itself is nowhere clearly delimited. The first three chapters, representing about one-quarter of the whole work, are devoted to analysis of stress, analysis of strain, and the stress-strain relation respectively. These chapters follow more or less classical lines, even to discussion of stress and strain quadrics, but the approach is almost entirely mathematical, and tensor notation is freely used. It may be remarked here that the mathematical background is heavily emphasized throughout, so that the title might well have been the " theory of elasticity for mathematicians " ; the mere engineer may perhaps feel that affine transformation, tensor fields, etc., scarcely pay their passage in the later chapters of application. Some few passages are likely to please neither mathematician nor engineer. For instance, the passage " the fact that the strains components e_{ij} cannot be prescribed arbitrarily can be seen from the following geometrical considerations ", leading after four pages of mathematics to the conditions of compatibility, makes the worst of both worlds. Whilst the passage " It is clear from physical considerations that one must demand that the functions u, v, w (the displacements) be single valued and continuous functions throughout the region occupied by the body " scarcely helps the engineer and cannot possibly satisfy the mathematician. The latter may be mollified later by a separate treatment of multiply connected regions ; but this latter passage is likely only to confuse the engineer, whose shaken confidence may finally disappear entirely in the section on uniqueness of solution.

After deploring the lack of uniformity of notation, the author adopts yet one more : the new notation aims at combining the best elements of all the others, but will uniformity be achieved in this way ? Moreover, in " Young's modulus and the ratio of Poisson " (*sic*), the author strikes an entirely new and surprising discord. A similar perversity is noticeable in the exclusion of aluminium alloys from a list of typical materials ; in the " demand that the resultant force and moment vanish " in one section, followed almost immediately by the dynamical equations ; in the strain energy function inextricably tangled with Hooke's law ; and in the statement that " the *strains* within the body are not susceptible to direct experimental verification ". The author's point would be made equally well by substituting " stresses " for " strains ", so why insult the radiologists ?

The one serious fault in these early chapters is the omission of all but passing reference to large displacements. The last section of Chapter II, on finite deformation, may give the impression that the two fields are synonymous, particularly because this section opens with the claim that " all the principal results of the classical theory of infinitesimal strain " have previously been derived.

The best features of these chapters are the insistence that crystalline solids cannot be more than statistically isotropic, unless the single crystals themselves are isotropic, and the reserved manner of statement of St. Venant's principle.

Chapter IV starts badly with a discussion of " engineer's theory of bending ", which leads virtually nowhere ; the true evangelist in this field should teach the engineer to remember the shear deflection, not tempt him to forget it. However, having found a firm footing in mathematics by identifying the torsion problem as Neumann's problem, the author canters off (" instead of solving a problem of Neumann, we can equally well solve a problem of Dirichlet ") into a long and very valuable digression on the uses of the complex variable and conformal mapping. The engineer who has to struggle with the mathematician's w- and z-planes may feel a savage desire to retort with a blue-print or two ; but this section is worth while, and the engineer should remember that the author is being as kind as he may be. On the other hand, the engineer may relieve his pent-up feelings on the section on curvilinear co-ordinates. Too high a price can be paid for generality, and this section is bought in the black market ; and surely curvilinear co-ordinates should be related to the complex variable !

At the end of the chapter the author returns to the " engineer's theory of bending ", only to fail spectacularly to retrieve his initial fault. By way of recompense there is an excellent appendix on the Schwarz-Cristoffel transformation (which, incidentally, is one of the few mathematical tools the student is not expected to have already mastered).

Chapter V deals with the problems of plane strain and plane stress, which are shown to be identical. Once more solutions are developed in terms of complex variables, and the treatment appears thorough ; but the going is hard, and the engineer might be encouraged by more remunerative results than examples on circular and elliptic contours. But at least the connection between complex variables and curvilinear co-ordinates is now made moderately clear.

Chapter VI, on the deflection of plates, deals extensively with the aeolotropic plate in deference to the technical importance of plywood and fibre-filled plastics ; in view of other omissions this inclusion seems odd. Moreover, the whole section wears (to the reviewer) an unreal air, because large deflections creep in only towards the end ; and a reference back to Chapter II for the basic formulae is not in the least helpful for the reason already noted. The mode of treatment of the large deflection problem and the example chosen afford but a poor distorted glimpse of this field.

Undoubtedly the chief value of the work as a whole lies in its applications of the complex variable. These methods deserve to be better known, and the present exposition should prove a useful reference. Yet the mathematician does not sell his goods particularly well, and many who could use these methods will not master them from this treatise. If this stricture be judged hard, let the engineer ponder the author's definition of a plate : " A cylinder is called a plate if its height is small compared with the linear dimensions of its cross-section " : and to think that the reviewer has meekly accepted criticism for speaking of the " slightly curved plate " !

H. L. COX.

Infra-red and Raman Spectra of Polyatomic Molecules. by GERHARD HERZBERG, Pp. iv + 632. (New York: D. Van Nostrand Co., 1945). $9.50.

In an important monograph published some years ago, Professor Herzberg gave us a thorough survey of the principles underlying the different kinds of spectrum associated with diatomic molecules, and their relevance to physical and chemical problems. A second similar volume had been planned to deal with the spectra of polyatomic molecules. As might have been expected, however, the rapid accumulation of experimental results, as well as the related theoretical background, made it necessary to divide the volume into two parts. The present book on infra-red and Raman spectra forms the first part, and a further volume is promised which will deal with the ultra-violet spectra.

Any reader of this book, not least the spectroscopic expert, will be deeply impressed by the vast amount of material collected by the author and by the clear way in which he has set it out. Several other recent monographs have dealt with either the infra-red or the Raman spectra, and many experimental and factual details can be found in them which are not included in the present book. However, the close relationships between the two types of spectrum make it desirable that, for many purposes at least, they should be considered together, and in this book the author has given us an admirable correlation of the data, as well as the most complete record of infra-red measurements yet published.

The infra-red spectra of even the simplest polyatomic molecules are far more complex than those of diatomic molecules. This arises because of the very marked increase in the number of possible molecular energy levels as the number of atoms in the molecule increases, and in their highly complex arrangement as regards magnitude. Elucidation of the spectra demands a detailed understanding of the energy states, and the rules governing spectral transitions between them. If the exact relationships are known between the values of the molecular energy levels on the one hand and the molecular dimensions, nuclear masses and electronic structures on the other, it may be possible to derive important structural information from the spectroscopic data. Even when the quantitative correlations are incomplete, useful qualitative information is often obtained.

The author has discussed all these matters exhaustively and systematically, and although in such a masterly presentation the reader's path must necessarily be hard, the end fully justifies the journey. The general reader may be surprised to find a substantial amount of ancillary apparatus here for use in the theoretical treatment. For example, there is a lucid, though concise, account of symmetry operations and point groups, and a detailed treatment of the mathematical technique for dealing with the harmonic vibrations of molecular models using symmetry or normal co-ordinates. As regards the symmetry properties, the arguments are illustrated by clear diagrams, and examples are given of molecules belonging to the different point groups. From the point of view of the physicist or chemist this is perhaps the most satisfactory and usable summary of this subject yet published. The methods given for calculating the normal vibration frequencies of a molecule on the basis of a given potential energy function illustrate not only the importance of the different types of force field which can be assumed but also the procedure for solving the secular equations which arise, and the simplification of such equations by the introduction of conditions of symmetry. The author has reviewed comprehensively all existing information obtained by applying such calculations to simple molecules whose frequencies of vibration are known. This makes it possible to assess in proper perspective the values obtained for the force constant of particular linkages in different molecules.

These discussions on normal vibrations lead naturally to the consideration of other related matters such as torsional motions and inversion doubling found with pyramidal structures like ammonia. The author also gives an informative account of the product rules of Teller and Redlich which relate the frequencies of a pair of isotopic molecules to each other independently of the particular potential energy function which may apply. In about a hundred closely packed pages full references are given to work on the vibrational spectra of all the important molecules yet studied, and there is also a critical discussion of the significant points which have emerged. This compilation alone is very valuable, but it also serves to bring out some of the special cases which require further examination.

There follows a detailed consideration of interactions between vibration and rotation, illustrated by the data for many vibration-rotation spectra. The important phenomenon of Coriolis interaction is discussed in some detail, although here the specialist may regard some of the interpretations as problematical. The remarkable increase in complexity of the rotational fine structure of vibration bands of asymmetrical molecules compared with linear molecules or symmetrical tops is illustrated both by the theoretical considerations of the energy levels and by the actual examples chosen. The latter serve to emphasize the point that even a small departure from the symmetrical top may lead to markedly increased complexity of the rotational fine structure. Most of the examples discussed in detail deal with resolved rotational fine structure. In reality, however, asymmetrical molecules are the majority, and most of them have moments of inertia which are too big to lead to such a resolved structure of the rotational branches. In such cases band contours are observed, which will be hybrids if the vibration concerned has components of change of electric moment along more than one of the axes of inertia. The question then arises whether these contours can be used as reliable guides in the assignment of vibrations to particular types, or even in the approximate determination of moments of inertia. The author has summarized briefly the theoretical computations of the appropriate band envelopes to be expected for different degrees of asymmetry, but in view of the growing number of experimental measurements in this connexion a fuller account might have been welcome.

There is a final chapter entitled ***Applications***, confined almost entirely to a polished account of the calculation of thermodynamic properties from spectroscopic data. The book concludes with an excellent bibliography and index.

Some may feel, after reading this book, that, admirable as it is, it still remains somewhat cold, rigidly austere and perhaps a little idealized in its treatment, for it deals to a large extent with the derivation of the energy levels of molecules having some degree of symmetry, and also having small moments of inertia and being capable of examination in the vapour state, where the rotational fine structure of the bands can be obtained. This criticism may be valid, and has some point in the sense that the majority of molecules in practice are asymmetrical rotators, have large moments of inertia, and are comparatively involatile. But the author's failure to deal with these molecules in the same exhaustive way is not at all detrimental to the book as it stands. The division is in some respects the division between physics and chemistry. In order to make full use of infra-red and Raman spectroscopy in chemistry, where the bigger and unsymmetrical molecules have the stage, an entirely different and more empirical approach may be required. Much indeed can be learnt in this way about intermolecular forces such as hydrogen bonding, and much can be done by the application of infra-red and Raman measurements to the question of structural diagnosis for special intramolecular groupings. Also the whole new field of infra-red analysis, both qualitative and quantitative, has recently been developed. A proper discussion of all these matters would have seriously lengthened the book, and in any case many of them are in a state of rapid growth and flux.

This book is a landmark for those directly interested in molecular spectroscopy. Others, both physicists and chemists, wishing to understand how infra-red and Raman spectra can be applied to the determination of molecular vibrations and the molecular structure of simple molecules will gain much from it. Both the author and the publishers are to be congratulated upon its production. H. W. T.

An Index of Mathematical Tables, by A. FLETCHER, J. C. P. MILLER and L. ROSENHEAD. Pp. vii + 450. (London: Scientific Computing Service, Ltd., 1946). 75*s*.

A worker who requires a figure from a mathematical table may desire either to work out how some other function varies, or may need just a particular value, for example the reciprocal of some number which occurs in his own calculations. In the former case, the need for an orderly table is evident ; in the latter, the user needs a table so that he may interpolate for his particular argument. The degree to which a table is necessary (as opposed to convenient) varies greatly. If one wants a reciprocal, it is to-day practically as easy to obtain it from a calculating machine as from a table ; if the value to be calculated is a wave-number from a wave-length, so that the reciprocal has also to be multiplied by a given factor, then the table of reciprocals is really of little value. A table giving the conversion direct is of value, but a table of reciprocals would need to be supplemented by a table of products or else both replaced by a table of logarithms. Again, when a calculation presents a sine as its answer, the argument is likely to be in radians, and numerical calculation from the formula demands either a table of sines with the argument in radians, or a table for converting radians to degrees. In navigation, on the other hand, the angles will nearly always be given in degrees, minutes and seconds, or occasionally in hours, minutes and seconds. One table of sines will serve all purposes if the user is prepared to carry out the conversion of the argument to a standard system, but it is undeniable that a great deal of time is saved in the aggregate if there are different tables for the different purposes.

Against these examples, of tables which are convenient but not indispensable, we may set those where the work really cannot be done if tables have not been prepared. Such cases arise where the needs are for factors of large numbers, or for values of Bessel or gamma functions or their inverses. The flow of heat or electric current in a cylinder involves Bessel functions, and if an experimentalist desired to calculate the constant of a material cylinder from observations on current flow in it, his task would be impracticable if he had to calculate the necessary functions instead of taking them from tables.

There are functions of an intermediate type, exemplified by $(1-x^2)^{\frac{1}{2}}$. This is not difficult to calculate by machine, especially if $1-x^2$ is factorized. It may also be obtained from tables of squares and square roots, or yet again from tables of sine and cosine (in any argument, provided it is the same for both). Yet many workers have thought it worth while to tabulate the function, and evidently a direct table must save much time if many values have to be worked out. Tallqvist seems to have been the first; he gave the function to 12 decimal places, at intervals of only 0·01, so that a good deal of interpolation would be called for. Fourteen years later, in 1920, T. Smith gave a four-figure table at intervals of 0·001 (0·0001 near unity); he was the first to give the table from $x=0{\cdot}7$ to $1{\cdot}0$, which enables the reader to obtain all values, by using the table as one of inverse functions over the rest of the range. Since 1920, according to the lists in the volume now under review, six other tables have been computed or published, to various degrees of accuracy. None of them follows Smith in his space-saving device.

Whichever type of table is needed by an investigator, whether the essential type or the convenient, it will be of no use to him unless he can learn quickly of its existence and location. The *Index* now under review (it will, I think, become known to mathematicians as the *Liverpool Index*) sets out to help in this problem. It consists of two main parts, of which the first is a list of functions, with the tables of each function set out in order. For each such table the *Index* sets out the accuracy (number of decimal or significant figures), range and interval of the argument, information about the means of interpolation (whether differences are tabulated and, if so, what orders, and whether proportional parts are shown), and finally the reference, in the form of the author's name and the year. In the second part, all authors are set out in alphabetical order, with their different publications by date, so that the references in the first part become precise by reference to this second part. In addition to the main parts, we should mention an interesting introduction by the two authors whose names stand first, another by Prof. Rosenhead (whose war work detached him from most of the work, after it had been planned), a publisher's preface by Dr. Comrie and a foreword by Prof. Hartree.

We learn in the preface that the authors do not believe they have attained completeness, and even that they could have approached nearer to that ideal if conditions had been different in the last six or seven years. This is no doubt true, but there are no striking omissions; the reviewer has compared the references in part 2 for the letters A and B with his own notes, and found far more in the *Index* than in his MS. The only converse cases are a publication in 1859 of a table by C. D'Aiguières, ***Tables sans fin donnant les résultats de la multiplication, de la division et de l'extraction des racines carrées et cubiques de tous les nombres imaginables***; J. Blater and A. Steinhauser, 1888, ***Multiplication and Division rendered rapid and easy by the calculation table, giving the 9 multiples of all numbers***; and J. Brown's ***Mathematical Tables*** (Edinburgh, 1830). None of these can be of any importance. I have no details of the French work; there are plenty of multiplication tables, and the collection by Brown, although it ran to at least five editions, was not a specially good one.

The need for the volume is really very great, and as far as can be judged without a year or so of trial, it has been extremely well done. That is, the choice of functions deliberately omitted is wise—there is no reason to suppose that there are important accidental omissions—and the information given seems to be accurate. The volume is really well produced, and for this it is clear that the mathematical world owes much to Dr. Comrie, who has given not only time and trouble, but has himself, as Scientific Computing Service, borne a very considerable expense (running to some thousands of pounds). To say that every scientific library needs the book is an understatement. It should, despite its cost, be bought by as many working mathematicians and mathematical physicists as possible.

J. H. A.